2011

Practical argumentation in a mixed-initiative framework

Chee Fon Chang
University of Wollongong

UNIVERSITY OF WOLLONGONG

Practical Argumentation in a Mixed-Initiative Framework

A thesis submitted in fulfillment of the
requirements for the award of the degree

Doctor of Philosophy

from

UNIVERSITY OF WOLLONGONG

by

Chee Fon Chang, B Com (Hons)

School of Computer Science and Software Engineering
December 2011

Dedicated to
My Family and Friends

Declaration

This is to certify that the work reported in this thesis was done by the author, unless otherwise referenced or acknowledged below. The document has not been submitted for qualifications at any other academic institution.

Chee Fon Chang, B Com (Hons)
December 14, 2011

Abstract

There exist many approaches to agent-based conflict resolution. Of particular interest is on approaches that adopt argumentation as their underlying conflict resolution machinery. One such approach is to view the argumentation process as an approach to attack, persuade an opponent or to defend one's belief. This dissertation proposes an abstract accrual argumentation framework that re-evaluates the treatment and utilisation of preference values within argumentation. We firstly present an incremental improvement on existing work to capture accrual of arguments within an abstract argumentation framework. Drawing from the incremental improvement, this dissertation highlights the importance of information source in argumentation and the effect on agent's decision making during argumentation. In most argumentation systems, the argument source plays a minimal role. We feel that ignoring this important attribute of human argumentation process reduces the capabilities of current argumentation systems. Secondly, this dissertation identifies the need for justification management in a setting where multi-agent performs negotiation or argumentation. An outcome-driven justification management framework is proposed in which traditional approaches in argumentation are modified to assist in the elicitation and management of justifications hence permitting the novel conception of mixed-initiative argumentation. Finally, the framework is also evaluated in the context of a clinical group decision support in medicine.

Acknowledgments

I would firstly like to thank my supervisor, Professor Aditya K. Ghose, for his guidance and advice throughout this long but interesting journey. His expertise in the area of Knowledge Representation & Reasoning has been most insightful.

I would also like to thank my fellow students in the Decision Systems Lab, with special mention to Peter Harvey for his comments on some of the material presented in this dissertation. Further thanks go to all the members of the Decision Systems Lab for creating an interesting and stimulating research environment.

Finally, I would like to thank my parents, for all their patience, understanding and support.

The following describes the interesting journey one takes when performing research and further highlights the difference between a teaching and a learning institute.
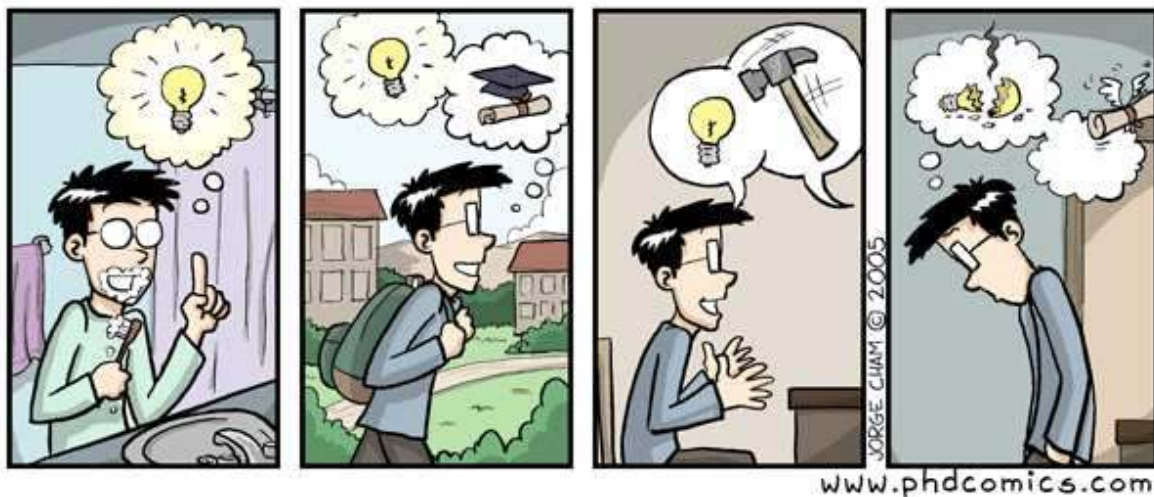
"Piled Higher and Deeper" by Jorge Cham www.phdcomics.com
Copyright ©1997-2005 Jorge Cham

List of Publications

- Graham Billiau, Chee-Fon Chang, Aditya Ghose and Alexis Andrew Miller. SBDO: A New Robust Approach to Dynamic Distributed Constraint Optimisation. In *Proceedings of the 13th International Conference on Principles and Practice of Multi-Agent Systems*, pages 11–26, November 2010, Kolkata India.

- Graham Billiau, Chee-Fon Chang and Aditya Ghose. Using Distributed Agents for Patient Scheduling. In *Proceedings of the 13th International Conference on Principles and Practice of Multi-Agent Systems (Revised Selected Papers)*, pages 551–560, November 2010, Kolkata India.

- Graham Billiau, Chee-Fon Chang, Alexis Andrew Miller and Aditya Ghose. Support-Based Distributed Optimisation: An Approach To Radiotherapy Patient Scheduling. In *Proceedings of the 8th HealthGrid conference*, pages 327–334, June 2010, Paris, France.

- Chee-Fon Chang, Aditya Ghose and Alexis Andrew Miller. Mixed-initiative argumentation: A framework for justification management in clinical group decision support. In *Proceedings of the Uses of Computational Argumentation, AAAI Fall Symposium*, pages 14–19, November 2009, Arlington Virginia, USA.

- Chee-Fon Chang, Alexis Andrew Miller and Aditya Ghose. Mixed-initiative argumentation: Group decision support in medicine. In *Proceedings of the 2nd International ICST Conference on Electronic Healthcare for the 21st century*, pages 43–50, October 2009, Istanbul, Turkey.

- Peter Harvey, Chee-Fon Chang and Aditya Ghose. Support-Based Distributed Search: A New Approach for Multiagent Constraint Processing. In *Argumentation in Multi-Agent Systems, Third International Workshop, ArgMAS 2006, Hakodate, Japan, May 8, 2006, Revised Selected and Invited Papers*, pages 91–106, 2007.

- Peter Harvey, Chee-Fon Chang and Aditya Ghose. Simple support based distributed search. In *Proceedings of the 19th Canadian Conference on Artificial Intelligence*, pages 159–170, June 7-9 2006, Quebec city, Canada.

- Chee-Fon Chang, Peter Harvey and Aditya Ghose. Source-Sensitive Argumentation System. In *Proceedings of the 8th International Conference on Enterprise Information Systems*, pages 39–46, 23-24 May 2006, Paphos, Cyprus.

- Chee-Fon Chang, Peter Harvey and Aditya Ghose. Combining Credibility In a Source-Sensitive Argumentation System. In *Proceedings of the 4th Hellenic Conference on Artificial Intelligence*, pages 478–481, 18-20 May 2006, Heraklio, Crete.

- Peter Harvey, Chee-Fon Chang, and Aditya Ghose. Support-based distributed search: a new approach for multiagent constraint processing. In *Proceedings of the 3rd International Workshop on Argumentation in Multi-Agent Systems*, pages 91–106, May 8 2006, Hakodate, Japan

- Peter Harvey, Chee-Fon Chang, and Aditya Ghose. Support-based distributed search: a new approach for multiagent constraint processing. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent systems*, pages 159–170, May 2006.

- Peter Harvey, Chee-Fon Chang, and Aditya Ghose. Practical application of support based distributed search. In *Proceedings of the 17th IEEE International Conference on Tools with AI*, pages 34-38, November 2005.

- Peter Harvey, Chee-Fon Chang, and Aditya Ghose. Support-based distributed search. In Amnon Meisels, editor, *Proceedings of the 6th International Workshop on Distributed Constraint Reasoning*, July - August 2005.

- Farzad Salim, Chee-Fon Chang, Aneesh Krishna, and Aditya Ghose. Towards Executabale Specification: Combining i* and AgentSpeak(L). In *Proceedings of the 17th International Conference on Software Engineering and Knowledge Engineering*, pages 739–742, 14-16 July 2005, Taipei, Taiwan, ROC.

Contents

List of Tables

List of Figures

List of Procedures

Part I

Background and Overview

1

Introduction And Motivation

"The aim of argument, or of discussion,
should not be victory, but progress."
– Joseph Joubert

1.1 Introduction

The underlying theme of this dissertation is on the theory of defeasible argumentation. As argued by Pollock in [110] and for the purpose of this discussion, the concepts behind defeasible reasoning as studied in Philosophy and non-monotonic reasoning in Artificial Intelligence are fundamentally equivalent. This dissertation develops upon and extends on these formalisation, therefore in this dissertation the terms used from both models are interchangeable.

In this chapter, we will begin with a brief discussion on the general concept of argumentation utilised throughout this dissertation. We put forward our perspective on the distinction

between multi-agent[1] argumentation and negotiation. Furthermore, we stipulate that the reasoning and justification properties of argumentation in multi-agent environments warrants further investigation. We will draw upon a series of examples to provide motivation for this research. The introductory chapter concludes with the research goals, contributions and an outline of the dissertation.

The study of arguments and argumentation is traditional to many disciplines. Although the notion of argumentation is common to most of these disciplines, there is still no consensus as to the "correct" meaning to the term [64]. The following highlights that the acceptability criteria within argumentation is subjective and is prone to persuasion. Hence, meeting an acceptability criteria is an accumulation of propositions justifying a standpoint.

> Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge. (van Eemeren et al. 1996 p5) [142]

Hence forth, it is our view that the focus of argumentation theory is *not* about parties or agents coming to an agreement through some concession by one or more parties. This notion is captured within the realm of negotiation. Negotiation is about collaboration where individual parties or agents trade positions on a given topic of discussion with the aim of coming to an agreement. These exchanges maybe achieved by deploying an underlying argument-theoretic procedure, such as those proposed by Jennings et al. in [76, 107, 136] where the focus is on the validity of communication protocol and negotiation procedures. Furthermore, negotiation works best when constraints (usually time or resources) exist, where parties participating in the negotiation are forced to compromise so that an agreement is reached within the given constraints. Parties participating in argumentation are not faced with such a dilemma.

Our venture into argumentation research is motivated by the fact that argumentation provides methods and techniques for addressing problems that have no definitive "correct" answers or solutions. It is our view that this phenomenon is due primarily to the subjective nature of an individual's acceptability criteria. Furthermore, we believe that an individual's acceptability criteria is a non-static measure, constantly changing during an argumentation process. In a multi-agent argumentation scenario, these acceptability criteria take on an even greater role, not only in the generation of solutions but also in justifying a selected or accepted solution.

---

[1]We will take the most general definition for the term "agent", hence do not specifically distinguish between human agents and software agents.

In the literature, there exists two distinct flavours to the theory of argumentation. One approach views arguments as logical proof and the argumentation process as a search for finding suitable logical statements. Another approach views the argumentation process as an approach to attack, persuade an opponent or to defend one's belief. Although, these two approaches are not mutually exclusive, in this dissertation, we have adopted the latter view in which argumentation theory is about persuasion. Argumentation theory is hence defined as the science of effective civil debate or dialogue, using rules of inference and logic, as applied in a *real world* setting. Therefore, argumentation theory in such a situation is not only concerned primarily with reaching "correct" conclusions through logical reasoning, starting with certain premises but also with the persuasive force of an argument. Given the flexibility resulting from the combination of individual acceptability criteria for the truth or falsity for any given assertion, sentence or argument, argumentation allows us to persuade others of our views on multiple facets. However, argumentation theory is *not* about logical *truth*, although logical truth maybe the resulting artefact of an argumentation process. As previously mentioned, the acceptability criteria of individuals participating in an argumentation process may vary and so does the perceived truth or falsity of an assertion, sentence or argument. Arguments may also be constructed from a logical representation language. The availability of well formed logical sentences is considered the fundamental to the construction of arguments, hence argumentation is not concerned with the complexity that maybe involved when constructing these logical sentences. Although not concerned with the construction and consistencies of the underlying logical sentences, at a meta-level, argumentation is aimed at reasoning about the semantics captured within these sentences. This allows argumentation the freedom to provide its own notion of consistency or meta-consistency, captured in the forms of *conflict*, *attack* or *defeat* relations. These notions of consistencies can be and are, in many situations, evidently weaker than that enforced in the logical language. Argumentation theory is therefore concerned with acceptability, and not necessarily any notion of truth or agreement. Given that this is the case, it is then inevitable that some notions of preference are required when dealing with individual acceptability criteria. Ultimately, this acceptability criteria is a reflection of an individual's preference on a set of arguments.

Studies on formal argumentation such as [5, 8, 28, 34] tend to focus on the representation of arguments, the structure of arguments and their interactions. The notion of preference is generally implicitly captured within either the *attack* or *defeat* relations. It is generally recognised ([6–11, 13, 22, 25, 49, 50, 77, 78, 100]) that this approach will not suffice. Recent studies such as [22, 49, 50, 77, 78] recognised that the acceptability of arguments is subjective, and is subjected to agreements between the participants of the argumentation exchange. Their notion of acceptability is derived from social values obtained from the participants and

audience.

Our approach differs from these approaches, in that we explicitly represent preferences as a set of abstract values. These values are not specifically embedded as part of the arguments and are utilised only as a tie-breaking mechanism. In effect, we have taken a modular approach in the utilisation of preference values to enhance the underlying argumentation system. This approach provides a greater degree of freedom, allowing us to capture interactions between arguments not possible in current abstract argumentation frameworks. Furthermore, the uniqueness of this dissertation is not only how the preferences are formulated or encoded but also how they are utilised for decision making and decision support. Firstly, we propose the use of these preference values to capture the notion of strength embedded within the *accrual* of arguments in an abstract argumentation framework. Secondly, we propose the use of this abstract preference-based accrual argumentation framework in argumentation-based evolution in a mixed-initiative fashion.

Research in this area of defeasible argumentation can be characterised by the studies of Pollock, Prakken & Sartor and Veheij. Most of these argumentation systems generalise to that of Dung [52]. The acceptability notion captured by Dung [52] is an implicit approach to the accrual of arguments, however this approach does not allow for the ranking of sets of acceptable arguments. One might argue that these rankings are outside the scope of an argumentation framework and reside with the selection function of the reasoner. However, we believe that when argumentation is deployed for group decision support, the ranking forms an intricate component of the decision making machinery. Furthermore, the ranking permits the arguments to play a more significant role as well as informing the reasoner as to which is the better choice.

Within argumentation, "accrual" generally refers to the grouping of arguments to support or refute a particular position. The grouping of arguments could, in some cases, strengthen a position and in others weaken them. Such a grouping of argument should be interpreted as: in the presence of *argument X* and *argument Y*, the *position Z* is supported. As highlighted by Vreeswijk [151], the accrual of reasons, or accrual of arguments is a phenomenon that is less well mastered or understood in argumentation. Vreeswijk posses the question of whether, or under what conditions, reasons should accrue, and if there are general principles behind the accrual of reasons. Prakken [122] highlighted that within the literature of argumentation, two distinct approaches have been proposed in the formalisation of accrual. The first approach, taken by Pollock [116] and Prakken & Sartor [123, 124] encodes the accrual by hand as a conditional with a conjunction of the accruing reasons in the antecedent. This approach is considered as a knowledge representation (KR) approach. The second approach, considered

as an inference approach, instead, regards accrual as a step in the inference process. In this approach, some form of aggregation is performed after all relevant arguments based on individual reasons have been constructed. A mechanism such as that taken by Krause et al. in [87], by Hage et al. [65, 149] and Verheij in [145, 147], is then used to decide between the conflicting sets of reasons. Research into accrual of arguments within an abstract argumentation framework such as that proposed by Dung [52] is limited. We believe that this warrants further and in depth study.

In [122], Prakken proposed three basic principles for accrual argumentation: accrual may decrease the strength of the arguments involved; arguments involved in an accrual cannot be considered individually; flawed arguments may not be accrued. We will take these as guiding principles when dealing with "accrual" of arguments. However, our argumentation framework will be rely heavily on preference values, and so must necessarily utilise different structures to those proposed by Prakken. We will instead develop on the framework used by Bench-Capon et al. [22, 49, 50] in their investigations into the effect of "social values" on argumentation, Kaci et al. [77] and Bourguet et al. [25] in their investigations into capturing multiple contexts or values within a preference-based argumentation framework. Bench-Capon et al. argued that the social values represent the norms of the society or community, and hence have greater importance. Although we will not make direct use of arguments promoting the concept of "social values" in this dissertation, we believe that the fundamental structure of the framework, with minor modifications, can potentially be utilised to handle accrual of arguments. Although the notion of accrual of arguments and the use of preference values within argumentation is not new, we believe that our approach is unique in that it is the first instance where these two concepts of accrual and preference values are united. Such a unification allows for the proposition of an abstract formalisation to accrual in argumentation. As we utilise Dung's argumentation framework, in which most existing argumentation systems and framework generalises to, our approach can be perceived as an initial attempt to unify existing approaches to accrual in argumentation.

Situated in a dynamic environment such as in the 'real-world', it is generally accepted that any knowledge-base will require revision, such that the reasoning machinery continues to generate output that is *socially* acceptable with respect to changes in the environment. An *argumentation-base*[2] is no different. Note that in the argumentation literature [125], revision on arguments are never performed. New arguments are constructed and added into the argumentation-base. Our approach is aligned with such views. Let us now consider two instances in which we believe revision of an argumentation-base is required.

---

[2]We coined the term argumentation-base to represent an encapsulation of all the assertions and associated relation required to perform argumentation-based reasoning.

Firstly, when constructing an argumentation system, it is conceivably a time-consuming and in most cases, an impractical exercise to extract from a domain expert all relevant knowledge. In fact, the simplest exercise of determining what is the relevant knowledge, and to what granularity or specificity should such knowledge be represented in the argumentation-base can be a challenging issue to address. Hence, given the effort required to perform this task, it is inadvisable to reconstruct an argumentation-base from first principles whenever the environment changes. Recent studies such as [16, 35] recognise the dynamic nature of an argumentation-base and the need for revision. However, these studies do not consider the varying degree of an individual's acceptability criteria, hence resulting in the need for the revision of the preference model represented within the argumentation system. Furthermore, these studies assumes the existence of machinery capable of extracting relevant knowledge from domain experts. We believe that this machinery is an intricate component of the revision exercise and hence should not be relegated to being second class citizen.

Secondly, argumentation systems are generally deployed to compute rationale/arguments for supporting decisions. In such a setting, the only reason for an individual to reject the generated arguments is because the reasoning is flawed. Such a situation is an artefact of three possible scenario. Firstly, the argumentation-base contains information that is incomplete or imprecise. Generally, this is not an issue. However, without a more complete argumentation-base, the reasoning machinery relies on assumptions which may not be aligned with the individual, hence, resulting in flawed conclusions. Secondly, there exists a possibility that the acceptability criteria of the individual is inaccurately modelled within the argumentation system. Without an accurate model of the acceptability criteria, the arguments that are considered acceptable by the reasoning machinery may indeed be unacceptable. Finally, it is conceivable that the argumentation-base contains information that is no longer aligned with the 'real-world'. In other words, the reality has changed or evolved. The logical resolution for all these situations is to revise the argumentation-base. Furthermore, current studies have made little consideration of the effect of performing iterative argumentation-base revision. We believe this is an important issue as changes to the argumentation-base should be performed minimally. We draw upon studies in the area of belief revision and merging such as those of Alchourrón et al. [3], Gärdenfors [61], Darwiche et al. [47], Konieczny et al. [85] and Meyer [98, 99] as guiding principle for the revision and expansion of the argumentation-base. This provides us with a notion of minimal change or minimal deviation from the previous argumentation-base.

Our approach differs from that of Cayrol et al. [35, 36] and Amgoud et al. [16]. Cayrol et al. addressing the issue of extension revision. Amgoud et al. considered the issue of revision of the argumentation theory without explicitly considering minimal changes. Both approaches

do not capture the use of preferences in the argumentation system nor the role preference values play in argumentation-base revision. Our approach focuses on the minimal revision of the argumentation-base rather then minimal revision of the generated extensions. Furthermore, we propose to utilise mixed-initiative interactions to facilitate argumentation-base revision. Using mixed-initiative interactions to achieve the desired argumentation theory change allows for incremental changes from the previous argumentation-base, reflecting the essences of iterative revision. This approach provides us with the ability to perform traceability on decisions as well as retrospective reasoning in an argumentation framework, hence the ability to perform decision quality assurance in an argumentation framework.

The next section provides the motivating basis for our thesis. It reflects common real-world situations in which we believe current formal argumentation systems should capture. This further reinforces the view that studies into argumentation theory are not only interesting but yields methods and techniques to address problems that are subjective in nature.

1.2 Motivation

Decision support via argumentation in any multi-agent community requires agents to determine the acceptability of arguments given to them based on their own knowledge base. Associated with these knowledge bases are sets of preferences unique to each agent. Furthermore, it requires the agent to revise knowledge captured in their argumentation theory to match that perceived from the "real-world". In this section, we will present a series of motivating examples to illustrate the journey in which this dissertation will take as well as illustrate the importance and non-trivial nature of this study. In the following motivation, we will ignore the actual truth or falsity of the assertion. We only require that the participating parties believe in their own assertions.

Firstly, we will present our arguments as to why the ability for an argumentation system to perform accrual of arguments should be considered independently to the existence of any conjunctive or disjunctive connectives available in the underlying logical language. This promotes the view that arguments should be considered as atomic in an argumentation system and the accrual of arguments distinct to any aggregative connective that might be provided by the underlying logical language.

Secondly, we will present a sequence of examples utilising preferences for the accrual of arguments and illustrating the importance of information sources during an argumentation

process.

Finally, we will introduce the notion of justification management in a mixed-initiative fashion, allowing for the evolution of an argumentation theory to match knowledge perceived from the "real-world".

1.2.1   Accrual of arguments

The following example is taken from Verheij [145, 146]. Consider two people are debating whether a particular person (Bill) should be incarcerated. In this situation, neither party is likely to withdraw their argument. Assume that you are presented with the following two arguments supporting the incarceration:

> *Bill has robbed someone, therefore he should be jailed.*
> *Bill has assaulted someone, therefore he should be jailed.*

In [151], Vreeswijk argued that the accrual of reasons, or accrual of arguments [117, 146] poses the questions of under what conditions should reasons accrue and whether there exists general principles behind the accrual of reasons. Vreeswijk maintained that accrual of reasons cannot be modelled in argumentation, citing the fact that rule bases of argumentation system are typically antecedent-incomplete and therefore contain insufficient information as to decide whether supports are strengthened or weakened among rules that share consequences. Conversely, once an argument is formed, the argument represents a specific application of rules within the rule base. Hence, without further information regarding the rule base of the argumentation system, it seems rather unwise to disassemble the arguments and perform accrual utilising the connectives in the underlying logical language. In doing so, this might change the original intent captured within the rule base and of the arguments. To illustrate this point, let us consider the two basic connectives: the conjunction and disjunction connectives. First, consider the use of the conjunction connective. We could rewrite the above arguments as:

> *Bill has robbed and assaulted someone, therefore he should be jailed.*

By merging the supports of the two arguments into one, the new argument would imply that Bill would have to commit both a robbery and an assault before he should be incarcerated. This argument can be easily defeated if an attack is mounted on either robbery or assault

whereas the in the previous setting, both arguments have to be defeated independently. By using the conjunction connective, the resulting defence condition would seem somewhat too strong. Evidently, this did not contribute to the increase in the concluding force. In fact, the reverse had occurred. Now, consider the use of the disjunction connective. We could rewrite the above argument as:

> *Bill has robbed or assaulted someone, therefore he should be jailed.*

On the surface, the rewrite seems to have maintained the original intention. However, we argue that the resulting defence condition is too weak. This argument is only defeated in the presence of an argument that stipulate that Bill did not both rob and assault anyone. Hence arguments such as: "*Bill did not rob someone, therefore he should not be jailed.*" or "*Bill did not assault someone, therefore he should not be jailed.*" considered pair-wised with the previous argument will not have sufficient strength to defeat it. Our point here is that accrual of arguments should be a process performed on the arguments and should not be performed at the underlying logical language once an argument has been constructed. This view of accrual of arguments therefore differs from that proposed by [117, 146]. Accrual performed at the underlying logical language, depict a revision of arguments. As observed by [125], in the event of new information, arguments are never revised, new arguments are generated. A revision of arguments points to an error in the argument construction process.

Furthermore, associated with each argument is the subjective notion of strength. This notion is not captured by the underlying logical language, hence the resulting rewritten argument could result in undesirable modification of this associated strength value. This is particularly true if the background knowledge (i.e. the rule base) for the generated argument is not available or incomplete. Furthermore, if we are to consider these arguments within an abstract argumentation framework, arguments are considered as primitive. This situation further emphasises the need for the ability to perform accrual on abstract arguments, hence, providing a more general principle behind the accrual of arguments. We can conclude that by trying to perform accrual of arguments at the level of the underlying logical language produces the following undesirable consequences:

1. If accrual is performed in the underlying logical language, it masks the interplay and interaction between arguments.

2. The conjunction of antecedent of an inference rule places undue constraints on the argument and hence weakens the argument.

3. The disjunction of antecedent of an inference weakens the rule. The consequence of this is a loss of information. It is often not clear which component of the antecedent is undefeated. Furthermore, topics that are unrelated maybe combined and this will hinder reasoning.

4. The manipulation of underlying logical representation once the arguments have been constructed may change the intended meaning of the argument.

The following examples will illustrate the unique nature of accrual in argumentation. The next example will highlight the importance of individual preferences in increasing the concluding force during the accrual of arguments.

1.2.2 The role of preferences in the accrual of arguments

Let us now expand on the previous example. Suppose you are *privileged*[3] to a debate between two people arguing whether a particular person (Bill) should be incarcerated. In this situation, one argues that he should because of a crime he has committed, while the other argues that he should not, because of his age. Assume you are presented with the following independently:

A_1:*Bill has assaulted someone, therefore he should be jailed.*
A_3:*Bill is a juvenile, therefore he should not go to jail.*

It is clear that these two arguments are in conflict and that from a simple examination of the arguments and the attack relation; it is not possible to determine which argument is stronger. So let us assume that we are sympathetic to the fact that Bill is a juvenile and our preference is not for Bill to be incarcerated, we say that the argument A_3 is stronger or more preferred. Let us now disregard the previous arguments and consider a new scenario presented to you with a different set of arguments:

A_2:*Bill has robbed someone, therefore he should be jailed.*
A_3:*Bill is a juvenile, therefore he should not go to jail.*

---

[3]The notion of privileged is especially important. It emphasises the association of argument with information ownership and that information is revealed to authorised individuals. In this instance, the reader is simply passing judgement on the presented argument utilising his or her own knowledge base to determine which arguments are defeated.

Again, if we assume that we are sympathetic to the fact that Bill is a juvenile and our preference is not for Bill to be incarcerated, hence A_3 is stronger or more preferred. However, if all the relevant arguments from both sets are to be considered:

A_1:*Bill has robbed someone, therefore he should be jailed.*
A_2:*Bill has assaulted someone, therefore he should be jailed.*
A_3:*Bill is a juvenile, therefore he should not go to jail.*

When the arguments from the both sets are accrued, conceivably the claims resulting from the accrual of A_1 and A_2 are stronger. In this situation, the A_3 might not be strong enough to *defend* against the attack of two arguments. This type of argument is referred to as *accrual* arguments, and has been extensively studied in [113, 122, 145]. Irrespective of our sympathy toward the fact that Bill is a juvenile, there exists a threshold in which the arguments against Bill would overwhelm our beliefs and force us to re-evaluate our preferences or position.

It is interesting to note that in the previous discussion, the reader has been called upon to perform the role of mediator, judge or jury. This demonstrates the importance of role separation. We have also demonstrated that when evaluating arguments, preferences influence individual acceptability criteria. This emphasises that the acceptability criteria is dynamic and is subjected to the influence from argument strength and preference. Furthermore, we have highlighted two distinct notions of strength: objective and subjective. The objective notion of strength reflects the logical process of the argumentation machinery while the subjective notion of strength reflects an individual's acceptability criteria. We have also demonstrated that the concluding force may increase when arguments supporting the same claim are accrued. This accrual process may force the re-evaluation of preferences and cause a revision of the acceptability criteria. Furthermore, it demonstrated that arguments are never evaluated independently of an individual's belief and preferences.

The following example will illustrate that accrual of arguments should not be performed irrespective of the source presenting the arguments and the role preferences play.

1.2.3   The role of information source in the accrual of arguments

Let us now further our argument by expanding on the previous example to demonstrate the issue of accrual from multiple sources. In this instance, we will associate the arguments with their sources. We will demonstrate that this information plays an important role during argumentation. Again, we are going to assume that you are privileged to a debate between

two people arguing whether a particular person (Bill) should be incarcerated. Previously, one argues that he should because of a crime he has committed, while the other argues that he should not because of his age. If the arguments are to be evaluated with consideration to the source of the arguments, the relative strength of each argument maybe be altered and hence the resulting defeat relation will vary. Given the two example scenario below:

Scenario 1
Tom: *Bill has robbed someone, therefore he should be jailed.*
Tom: *Bill has assaulted someone, therefore he should be jailed.*
Dick: *Bill is a juvenile, therefore he should not go to jail.*

Scenario 2
Tom: *Bill has robbed someone, therefore he should be jailed.*
Harry: *Bill has assaulted someone, therefore he should be jailed.*
Dick: *Bill is a juvenile, therefore he should not go to jail.*

Although the arguments presented in the two scenarios are the same, the person, agent or source presenting the arguments differs. We believe that this difference will affect the acceptability of the arguments as some sources maybe more preferred than others.

In example 1.2.2, we have accepted that the accrual of the first two arguments defeat the third. However, if we believed that *Dick* is a more trustworthy source of information, would his argument be easily defeated? How many more arguments from *Tom* would one require before changing one's acceptability criteria? Would the introduction of a different source, say *Harry* influence the decision? It is conceivable that the one source of information is more trustworthy relatively to another, hence an individual's preferences will reflect such bias. We argue that the arguments' source performs an important role in determining the acceptability of the argument. This example demonstrates that preferences are usually informed by knowledge external to the argumentation system. This knowledge could be associated with an individual's belief such as credibility of the participants.

The following example will illustrate the importance of preferences informed by social perception.

1.2.4   The role of repeating sources in the accrual of arguments

With the introduction of the notion of person, agents or sources, an additional question arises. For this example, let us assume that all sources are equally preferred. Consider the following example:

$Person_1$: *Bill has robbed someone, therefore he should be jailed.*
$Person_2$: *Bill has robbed someone, therefore he should be jailed.*
$Person_3$: *Bill has robbed someone, therefore he should be jailed.*
$$\vdots$$
$Person_n$: *Bill has robbed someone, therefore he should be jailed.*
$Person_m$: *Bill is a juvenile, therefore he should not go to jail.*

It is clear that in human argumentation, a set of logically identical arguments with different supporters are stronger than an argument with only one supporter[4]. This is a phenomenon seen in legal debate and typical elections. However, in most artificial argumentation systems, a set of repeated arguments either hold no additional weight, or are explicitly disallowed. We argue that repeating arguments, if provided from different sources, should be considered as distinct arguments and hence should strengthen the claim of an argument. This example demonstrates that preferences can be informed by knowledge external to the argumentation system to reflect behaviours such as social value or perception of a community.

The final example illustrates the use of argumentation for decision making in a dynamic environment where information is incomplete. In such a situation, revision of the preferences as well as the underlying argumentation theory are required. The management of argument justification over a sequence of decisions can assist in consistency management.

1.2.5   Justification management in argumentation

In this section, we will illustrate the need for justification management in a multi-agent argumentation setting. Let us consider the previous example 1.2.2. In the example, we performed a sequence of three decisions. In the first and second argumentation exchanges, we concluded that Bill should not go to jail. We came to the same conclusion for both situations because we were sympathetic to the fact that Bill is a juvenile and our preference is not for Bill to go to jail. However, in the third argumentation exchange, we decided that

---

[4]By supporter, we simply mean sources

the crimes that Bill have committed out rules our preference for Bill not to go to jail. This suggests a change or revision of our preference is required.

To illustrate our point, let us consider as a point of departure, an argumentation system with an empty argumentation background knowledge-base. In the first argumentation exchange, we were presented with two arguments:

> A_1:*Bill has assaulted someone, therefore he should be jailed.*
> A_3:*Bill is a juvenile, therefore he should not go to jail.*

In this situation, we concluded that Bill should not go to jail. By deciding that Bill should not go to jail, we have indicated our preference over arguments. Assume that these two arguments are added to the argumentation system. For the argumentation system to come to the same conclusion, a preference rule stating that A_3 is more preferred needs to be added. Now let us consider the second argumentation exchange where we were presented with the following two more arguments:

> A_2:*Bill has robbed someone, therefore he should be jailed.*
> A_3:*Bill is a juvenile, therefore he should not go to jail.*

In this case, we again asserted that A_3 is more preferred. Assume that the arguments are added to the argumentation system. No new preferences rule changes are required as the existing rule will generate the expected decision. However, in the final argumentation exchange, we were presented with the following arguments:

> A_1:*Bill has robbed someone, therefore he should be jailed.*
> A_2:*Bill has assaulted someone, therefore he should be jailed.*
> A_3:*Bill is a juvenile, therefore he should not go to jail.*

In this case, we concluded that A_3 is defeated. In this situation, the existing argumentation system will not yield the expected decision. Hence, existing rules will need to be modified and additional rules will need to be added. This can be viewed as an approach to check the consistency of an outcome over a sequence of decisions. This example illustrates that an argumentation system is not a static entity. The system needs to evolve as past decision rules may require revision and/or a new rule added to reflect the changes in the "real-world". One could find analogous in the legal system where once a precedence is set, subsequent

judgement should be consistent with that precedent. However, should a precedent be violated, additional facts and rules are provided to distinguish the new judgement from the past. This example also illustrated the need for an argumentation system to perform justification interaction so that it can evolve through use.

Existing literature in the area of accrual argumentation only consider how arguments should be accrued to support or refute a particular position. The question of whether, or under what conditions, reasons should accrue is not well studied. Furthermore, the management of justification within an argumentation framework is not well studied. Therefore, issues highlighted in examples 1.2.3, 1.2.4 and 1.2.5 are not currently addressed by any abstract accrual argumentation frameworks. By associating source to arguments, this provides the conditional and rationale required for a reasoner to determine when and where accrual should be performed. This further illustrates the gap within the area of accrual argumentation and support for further research.

1.3 Research Goals and Contribution

The thrust for this research is inspired by the fact that current available models of argumentation do not satisfactorily capture accrual of arguments and the utilisation of preferences. Our point of departure, although certainly not new, remains a valuable insight despite the abundance of presented argumentation models. Our aims are to firstly demonstrate that the study of argumentation is an interesting venture, which provides methods and techniques for addressing "real-world" problems. Secondly, to illustrate the importance of preferences in the accrual of argumentation and allowing this use of preferences to redefine the notion of acceptability criteria. Finally, to illustrate that an argumentation system should not be static but rather constantly evolving. Such evolution is not performed independently but depends on the previous iteration. Highlighting the need for argumentation-base iterative revision.

The contribution of this dissertation is then threefold. Firstly, it advances the state-of-art in the accrual of arguments in an abstract argumentation framework by the use of preference values. Following [22, 49, 50], we argue that persuasion in such cases where it is impossible to determine conclusively that either party is wrong, relies on a recognition that the acceptability of an argument depends on the individual's preference values and that the determination of which argument is more acceptable is subjective to the audience. Although the notion of accrual of arguments and the use of preference within argumentation is not new, we believe that our approach is unique in that firstly, it is the first instance where these two

concepts are unified. Furthermore, we have taken a distinctly unique approach by proposing the unification in an abstract preference-based accrual argumentation framework ($PAAF$).

Secondly, it identifies ways in which traditional approaches to argumentation can be modified to perform justification management in decision support where arguments and the process of arguing provides the justifications for decisions. The use of argumentation theory in group decision support has grown in significance in the past decade. When arguments are utilised to support decisions, they form the rationale or justifications in which the decision is based. However, these justifications mirror the "real world" knowledge and preferences of participating agents, hence any argumentation system is required to evolve as the environment it sits in changes. To perform such a task, the revision to an argumentation theory over a sequence of (argumentation-based) decisions is required. These revisions can potentially assist in the management of decision justification. We propose the use of an outcome-driven decision rationale management framework that permits a novel conception of the mixed-initiative argumentation Framework ($MIAF$)

Finally, we deploy these theoretical concepts in two distinct instances. In the first instance, we utilise the preference-based accrual argumentation framework in a generic multi-agent argumentation situation. The preferences-based accrual argumentation framework is deployed such that preferences are utilised to associate arguments with sources. As such, associating sources (in other words agents) to arguments highlights the ability for an argumentation system to capture intuitions such as information ownership, degree of reliability, credibility or trust.

In the second instance, the mixed-initiative argumentation framework is used to provide solutions in addressing issues within clinical decision support. The mixed-initiative argumentation approach is utilised for the elicitation of decision rules, evolution and maintenance of the argumentation-base, hence providing a model of argumentation that can be used to support decision making and more importantly the management of justifications over a sequence of decisions. Furthermore, it highlights the potential for argumentation to address decision quality assurance issues. The quality assurance of decisions entails the retention of past decision justification such that retrospective analysis can be performed on decisions to assessed the "correctness" of past decisions. By performing such retrospective analysis, a level of assurance can be maintain on future decisions.

1.4   Outline of the Dissertation

The structure of this dissertation is divided into three major parts. Part one consists of introduction, background and the foundation overview of the abstract argumentation framework (AF) proposed by Dung in [52], the preference-based argumentation framework (PAF) proposed by Bourguet in [25] and the value-based argumentation framework (VAF) proposed by Bench-Capon in [22].

In chapter 2, we present the background literature and concepts drawn upon in this dissertation. Furthermore, we highlight details of two foundational frameworks from which the theoretical framework presented in this dissertation drew inspiration as well as utilised as fundamental building blocks.

Part two contains the theoretical and application components of this dissertation. In chapter 3, we introduces the abstract preference-based accrual argumentation framework ($PAAF$). Within this chapter, we will motivate and propose approaches to capturing accrual by the use of abstract preferences. Using examples we highlight the intricacy and interplay as well as the properties of the framework.

In chapter 4, we introduces the mixed-initiative argumentation framework ($MIAF$). Within this chapter, we will firstly motivate the need for such a mixed-initiative argumentation framework and the use of the framework to perform decision support with the focus on decision justification and argumentation theory change. Furthermore, we propose a collection of procedures to perform minimal modification to the argumentation theory as well as a set of postulates to govern such argumentation theory revision.

In chapter 5, we present two distinct applications. Firstly, we present an application of the preference-based accrual argumentation framework. Within this section, we firstly describe the problem source-sensitive argumentation system ($SSAS$) addresses. We provide the vocabulary and machinery to perform source-sensitive argumentation. We conclude the section with a discussion on other areas of applications. Secondly, We present a practical application of the mixed-initiative argumentation framework within the domain of clinical decision support for oncology. Within this section, we firstly describe the problem in which our tool ("Just-Clinical") aims to address. We then proceed with describing the use of the tool. We conclude the section with a discussion on some issues and future applications.

Part three contains the conclusion, summary of results and some directions for further research.

2

Background and Overview

"Silence is one of the hardest arguments to refute."
– Josh Billings

2.1 Introduction

In chapter 1, we introduced and motivated our venture into argumentation. In order to truly appreciate the contribution made by this dissertation, it is important to first gain an understanding of the current state-of-the-art. In this chapter, we will introduce the existing literature to position our approach. This dissertation draws upon three distinct paradigm: argumentation theory, mixed-initiative interaction and belief theory change. The inclusion of literature from argumentation theory and mixed-initiative interaction is self explanatory. However, the relationship between these two paradigms and belief theory change may require some further elaboration. From the chapter 1, we highlighted our intent to marry argumentation theory and mixed-initiative interaction to produce an outcome-driven decision

rationale management framework. Any such outcome-driven decision rationale management framework necessarily requires modifications of the underlying argumentation theory during the rationale justification phrase. The AGM framework [3] within the belief theory change literature provides an important basis to understanding theory change. Inspired by the AGM approach, we can discuss certain properties that should govern argumentation theory change. Drawing analogues from the belief theory change literature allows for an elegant approach in prescribing the expected change on the argumentation theory during the coarse of mixed-initiative argumentation process. To properly appreciate the proposed properties for argumentation theory change, we believe it is beneficial for belief theory change literature (specifically the AGM framework) to be included in this chapter. We will provide an overview of each of these three paradigms in this chapter. However, the bulk of our background review will focus on argumentation theory since that is the main thrust of this dissertation.

We will focus on the construction of the Abstract Argumentation Framework (AF) proposed by Dung [52], the Preference-based Abstract Argumentation Framework proposed by Bourguet et al. [25] and the Value-based Abstract Argumentation Framework (VAF) proposed by Bench-Capon in [22]. These three argumentation frameworks are of particular interest as they provide the basis for our proposed argumentation framework.

2.2 Argumentation Frameworks and Systems

Several approaches such as default logics [19, 94–97, 101, 129] have been proposed for formalising non-monotonic reasoning. Argumentation provides an alternate perspective to the formalisation of non-monotonic reasoning or defeasible reasoning. In argumentation, a conclusion is accepted or withdrawn based on the interplay between the supporting and attacking arguments. This interplay relies on the relative as well as global strength of these arguments hence determining whether these arguments can be attacked and defeated by others. The notion of strength is usually implicitly defined within the argumentation system. This approach has been characterised as *defeasible argumentation*[1] and has raised a significant amount of interest after the initial work of Loui [90] and Pollock [110]. As by highlighted by Kraus et al. [86], the main purpose of these logics utilised in argumentation is to construct "defeasible proofs". These proofs are represented by arguments and ordered by relations placed on them, hence, expressing differences in conclusive force. In [138], Simari et al. treat arguments as

---

[1]A more comprehensive view of logics for defeasible argumentation can be found in [125]. Another survey on this topic, including a historical account of argumentation and defeasibility, can be found in [41].

prima facie proofs that may make use of assertions that one sentence is a (defeasible) reason for another.

In [120], Prakken proposed a generic conceptual framework underling the majority of the existing defeasibile argumentation systems. This generic framework consist of five components:

- A logical language with an associated logical consequence.

- A definition of an argument.

- A definition of conflict.

- A binary relation usually called a defeat relation.

- A definition of the status of an argument.

Note that these components may not be always be explicitly defined and that the terminology used to designate them may also vary between argumentation systems. The use of the logical language is to provide the necessary consequence relationship between the premise and conclusions. Utilising the logical language, an argument can be constructed. As such, an argument corresponds to a proof in the underlying logic.

Conflict can be viewed as disagreement or attacks between arguments. Conflict can be symmetrical and non-symmetrical. In [125], Prakken et al. highlighted three forms of attack rebuttal, assumption attack and undercut. Where rebuttal is described as arguments with contradicting conclusions. As such, rebuttals are a form of symmetrical conflict. Assumption attack and undercut are forms of non-symmetrical conflict. Assumption attack is expressed as a contradiction between facts and assumption. Undercut is when one argument challenges the rule of inference of the other argument [125]. Furthermore, Prakken et. al highlighted notions of direct and indirect attacks, where indirect attack is directed at sub-conclusions of an argument. It is interesting to note that the definition of undercut and assumption attack is not universality applied as in some argumentation systems such as [23, 106], indirect attacks are defined as undercut and in other systems [16–18], indirect attacks are defined as assumption attack. The notion of conflict or attack does not encompass any form of evaluation, hence defeat is utilised to express successful attacks. The status of an argument depends on the interaction within the whole set of arguments. Determining that status can be performed either declaratively, by defining a class of acceptable arguments; or procedurally, by constructing proof-theoretical machinery for determining whether an argument is in the class of acceptable arguments.

In [90,91], Loui represented a system of argumentation where defeat among arguments is defined recursively in terms of inference, specificity, directness and evidence. In addition, Loui developed an implementation for this rule system to compute defeat among arguments [91]. Lin and Shoham [89] developed an argument system that captures some well-known non-monotonic logics. However, in their system, they do not have logical a hierarchy among arguments, hence, it is not possible to determine which argument is undefeated.

Systems such as OSCAR, by Pollock [110–115] developed the argumentation system that can reason with suppositional arguments. Nute [104, 105] developed the LRD system and introduced so-called top-rules. Adjudication among competing arguments is performed via these top-rules. An argument defeats another if and only if the antecedent of the top-rule of the first argument is strictly more specific than the antecedent of the top-rule of the second.

In [150], Vreeswijk presented a critique of existing argumentation systems. He presented an abstract argumentation system where the basic notions of argumentation are well-defined but did not attempted to prescribe how argumentation should be performed, such as what arguments are in force or how defeasible information should be manipulated.

Dung [52] presented an abstract argumentation framework. The conceptual sketch provided by Prakken [120] is also in line with Dung's [52] view. Dung highlighted that every argumentation system consists of two essential parts: an Argument Generation Unit (AGU) and an Argumentation Processing Unit (APU). The AGU is used for generating arguments and the APU is used for deciding whether an argument is acceptable. In [52], Dung argues that logic programming and non-monotonic reasoning are types of argumentation which can be formalised in an abstract framework. Using a method for generating meta-interpreters for argumentation systems, Dung illustrated that argumentation can be seen as logic programming. This approach is illustrated below:

- The AGU specifies the attack (conflict) relationships between arguments. In [52], these relations are considered to be primitive and represented in terms of a binary predicate *attack*: if an argument α attacks an argument β, this is thus expressed as $attack(\alpha, \beta)$

- The APU is a negation as failure logic program consisting of the following two clauses that determines whether an argument α is acceptable:

$$acceptable(\alpha) \leftarrow \neg defeat(\alpha)$$
$$defeat(\alpha) \leftarrow attack(\beta, \alpha) \wedge acceptable(\beta)$$

  Intuitively, an argument is acceptable if it cannot be shown to be defeated, i.e. if there is no acceptable argument that defeats it. This captures the idea that an argument α

can be attacked by another argument, which in its turn may also be attacked by a third argument, therefore restoring (reinstating) the validity of α, but does not capture the distinction between justified and defensible arguments mention previously.

In the current landscape, authors such as Pollock [116], Verheij [145–149] and Prakken [121, 122] specifically performed studies into the notion of accrual argumentation where is most other researchers dismissed the notion as the use of conjunction within the underlying language. As a result there exists limited studies performed in accrual argumentation and the majority of the studies focuses on argumentation schemas and rules.

Verheij [145–147, 149] combines ideas of Lin & Shoham [89] and Vreeswijk [150] on the structure of arguments with Pollock's partial status assignments [116] into a formalism called CumulA. One of the result is the introduction of a new type of argument called coordinated argument, which combines two arguments for the same conclusion. With coordinated arguments Verheij aims to capture the accrual of arguments.This highlights that an investigation into an over arching abstract framework that allows for the capture for accrual is warranted.

The role of preferences in argumentation has also been mostly ignored. Again limited studies addresses the significance preferences have on deciding a "winner". These two issues are fundamental when argumentation is utilised as the formal basis for resolving conflict in agent negotiation. Studies such as [6–11, 13, 22, 25, 49, 50, 77, 78, 100] recognised the important role preference plays in argumentation. Recent studies such as [22, 49, 50, 77, 78] recognised that the acceptability of arguments is subjective, and is subjected to agreements between the participants of the argumentation exchange. Their notion of acceptability is derived from social values obtained from the participants and audience.

In next three sections, we will introduce the Abstract Argumentation Framework (AF) proposed by Dung in [52], the Preference-based Argumentation Abstract Framework (PAF) proposed by Bourguet et al. [25] and the Value-based Argumentation Framework (VAF) proposed by Bench-Capon in [22] as the foundation preliminaries to our proposal. One of the central themes of this dissertation is to emphasise the applicability of preferences in practical reasoning using argumentation theory. As discussed in the introduction chapter, our interest in argumentation theory does not lie solely on argument representation schemas and the establishment of soundness, but in situations where persuasion, defence or attack of one's viewpoint are the key motivation.

2.2.1 Abstract Argumentation Framework

One of the most influential contributions to the study of abstract argumentation systems is that of Dung [52] and the later extension of Bondarenko, et al. [24, 53] developed in the mid 1990s and early 2000. Our interest focuses on the structural construct and the notion of a preferred extension, hence we will not duplicate this work in its entirety. Interested readers are directed to [24, 52, 53] for the details. We will take, as the point of departure, the abstract argumentation framework (AF) proposed by Dung [52]. AF is concerned with capturing argumentation at an abstract level and showing that at this level of abstract, several systems of non-monotonic reasoning can be represented. Central to the AF is the notion of admissibility. Once a admissibility is established, several different types of semantics can be captured in the abstract framework. An abstract argumentation framework is defined as:

Definition 2.2.1 (Argumentation Framework [52]). *An Abstract Argumentation Framework (AF) is a pair:*

$$AF = \langle AR, attacks_{AF} \rangle$$

where

- *AR is a set of arguments.*

- *$attacks_{AF}$ is a binary relations on AR, i.e. $attacks_{AF} \subseteq AR \times AR$.*

For readability, we will denote $attacks_{AF}(\alpha, \beta)$ to mean α attacks β. We also say that a set of arguments S attacks an argument β if β is attacked by an argument in S.

Let us consider the motivating example 1.2.1.

Example 2.2.1. *Assume that $\{\alpha, \beta, \gamma\}$ are arguments representing "Bill is a juvenile; therefore he should not go to jail", "Bill has assaulted someone, therefore he should be jailed" and "Bill has robbed someone, therefore he should be jailed" respectively. The $attacks_{AF}$ relationship capturing the interplay between the arguments are represented as $\{attacks_{AF}(\alpha, \beta), attacks_{AF}(\beta, \alpha), attacks_{AF}(\alpha, \gamma)$ and $attacks_{AF}(\gamma, \alpha)\}$. Table 2.1 presents a summary of this discussion.*

Futhermore, utilising a digraph, we can illustrate the interaction between the arguments by representing arguments as labelled vertices and the attack relation as directed edges. Hence $attacks_{AF}(\alpha, \beta)$ is represented with a directed edge from the vertex α to β (see figure 2.1).

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{AF}$	$attacks_{AF}(\alpha, \beta), attacks_{AF}(\beta, \alpha)$ $attacks_{AF}(\gamma, \alpha), attacks_{AF}(\alpha, \gamma)$

AR is the set of arguments consisting of $\{\alpha, \beta, \gamma\}$. $attacks_{AF}$ is the set of binary relation on AR capturing the attack relation between the arguments in AR.

Table 2.1: Summary of Example 2.2.1
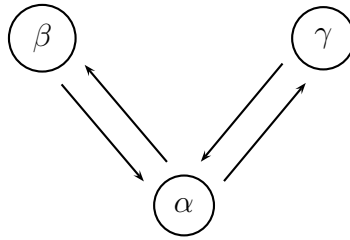
The arguments are represented by the three vertices α, β and γ. The directed edges represents the attack relations $attack(\alpha, \beta)$, $attack(\beta, \alpha)$, $attack(\alpha, \gamma)$ and $attack(\gamma, \alpha)$. Hence, the illustration can be interpreted as the argument α is attacking arguments β and γ while both arguments β and γ are attacking the argument α.

Figure 2.1: Graphical Illustration of Example 2.2.1

Given that there is no restriction placed on the set of arguments AR and the $attacks_{AF}$ relation, an AF alone does not encapsulate any notion of consistency. This is provided separately via a weaker notion, introduced as conflict-free. Conflict within a set of arguments can be described as the inability for two or more arguments to co-exists in the same set. Furthermore, conflict between sets of arguments can be captured as attacks between elements in the sets. The notion of conflict-free is then the lack of conflict between arguments within a set and is defined as:

Definition 2.2.2 (Conflict-free [52]). *Let $AF = \langle AR, attacks_{AF} \rangle$, a set of arguments $S \subseteq AR$ is said to be* conflict-free *if there are no arguments $\alpha, \beta \in S$ such that $attacks_{AF}(\alpha, \beta)$.*

Given the definition of conflict-free, let us consider the following example.

Example 2.2.2. *Continuing from example 2.2.1, the following conflict-free set exists: empty-set, $\{\alpha\}, \{\beta\}, \{\gamma\}$ and $\{\beta, \gamma\}$. Table 2.2 presents a summary of this discussion.*

We will now introduce the notion of acceptability. In a real world setting, an argument α is acceptable if an individual finds no reason to dispute α. In other words, the individual is able to defend α against any attacks. Hence, we say that an argument α, is acceptable to a set of arguments if the set defends α from attacking arguments that is not in the set. This notion

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{AF}$	$attacks_{AF}(\alpha, \beta), attacks_{AF}(\beta, \alpha)$ $attacks_{AF}(\gamma, \alpha), attacks_{AF}(\alpha, \gamma)$
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$

The empty-set, $\{\alpha\}, \{\beta\}, \{\gamma\}$ and $\{\beta, \gamma\}$ are conflict-free sets. The conflict-free sets can be computed by taking the power set of AR and eliminating all subsets of $\mathcal{P}(AR)$ that intersects with elements in $attacks_{AF}$.

Table 2.2: Summary of Example 2.2.2

of *defense*, or *reinstatement* of arguments is the basis for several of the semantics extensions presented in [22, 24, 25, 36, 52, 150]. Furthermore, given the notion of conflict-free and acceptability, we can construct a notion of admissibility. It is reasonable to assume that the acceptance of any set of arguments to an individual only occurs if the set of arguments is conflict-free and defends itself against all attacks. Hence we can say that a set of arguments is admissible if it is firstly conflict-free and that it defends itself from all external attacks. Therefore, acceptability and admissibility are defined as:

Definition 2.2.3 (Acceptable, Admissible [52]). *Let* $AF = \langle AR, attacks_{AF} \rangle$,

1. *An argument* $\alpha \in AR$ *is* acceptable *with respect to a set* $S \subseteq AR$ *of arguments if and only if for each argument* $\beta \in AR$: *if* $attacks_{AF}(\beta, \alpha)$ *then* β *is attacked by* S.[2]

2. *A conflict-free set of arguments* $S \subseteq AR$ *is* admissible *if and only if each argument in* S *is acceptable with respect to* S.

Given the definition of acceptability and admissibility, let us consider the following example.

Example 2.2.3. *Continuing from example 2.2.2, the following admissible set exists: empty-set,* $\{\alpha\}, \{\beta\}, \{\gamma\}$ *and* $\{\beta, \gamma\}$. *Table 2.3 presents a summary of this discussion.*

We will now describe the semantics for the AF. In most reasoning systems, there exist two basic semantics: credulous and sceptical. This is also the case for AF. Dung [52] defined several semantics such as preferred extension, stable extension semantics, fix-point semantics, grounded semantics as well as classifying the semantics into the two categories of credulous and sceptical semantics. For the purpose of this dissertation, we will focus solely on one of the credulous semantics: preferred extension. A set of arguments is considered a

---

[2]*This definition of acceptability by Dung does not require a set of arguments* S *to be conflict-free nor does it require that* S *not attack the acceptable argument* α.

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{AF}$	$attacks_{AF}(\alpha, \beta), attacks_{AF}(\beta, \alpha)$ $attacks_{AF}(\gamma, \alpha), attacks_{AF}(\alpha, \gamma)$
$conflict\text{-}free$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
$admissible$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$

The empty-set, $\{\alpha\}$, $\{\beta\}$, $\{\gamma\}$ and $\{\beta, \gamma\}$ are admissible sets. The set of admissible sets can be computed by filtering the set of conflict-free sets. For each conflict-free set S, firstly determine all arguments in AR that attack S. Secondly, determine the arguments in AR that is attacked by the arguments in S. The conflict-free set S is admissible if the set difference between the arguments attacking S and the arguments attacked by S is an empty-set. In other words, all arguments attacking S are attacked by some arguments in S.

Table 2.3: Summary of Example 2.2.3

preferred extension if it is a maximal defendable set of arguments. The notion of a preferred extension is defined as:

Definition 2.2.4 (Preferred extension [52]). *Let* $AF = \langle AR, attacks_{AF} \rangle$, *a preferred extension* *of an argumentation framework* AF *is a maximal (with respect to set inclusion) admissible set of* AF.

Given the definition of a preferred extension, let us consider the following example.

Example 2.2.4. *Continuing from example 2.2.3, the following preferred extension exists:* $\{\alpha\}$ *and* $\{\beta, \gamma\}$. *Table 2.4 presents a summary of this discussion.*

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{AF}$	$attacks_{AF}(\alpha, \beta), attacks_{AF}(\beta, \alpha)$ $attacks_{AF}(\gamma, \alpha), attacks_{AF}(\alpha, \gamma)$
$conflict\text{-}free$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
$admissible$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
$preferred$	$\{\alpha\}, \{\beta, \gamma\}$

The sets $\{\alpha\}$ and $\{\beta, \gamma\}$ are the preferred extensions. The preferred extension can be computed by filtering the set of admissible sets. For each admissible set, determine if the set can be extended by including an argument from AR while maintaining the admissibility condition. If the admissible set can be extended then it is not a preferred extension.

Table 2.4: Summary of Example 2.2.4

2.2.2 Preference-Based Abstract Argumentation Framework

The use of preferences within argumentation is not new [6–9, 11, 13, 22, 49, 50, 78]. In [25], Bourguet, et al. highlighted that the arguments captured in the abstract argumentation framework (AF) proposed by Dung [52] are assumed to have equal strength. In an attempt to

present a unified model for capturing preferences and ordering on arguments, Bourguet et al. [25] proposed the preference-based abstract argumentation framework (PAF). This unified framework subsumes systems/frameworks such as [9, 10, 77, 78] and allows for arguments to be associated with a (partial or total) pre-order preference relation hence capturing a notions of strengths, values as well as context for any given argument. Again, our interest focuses on the structural construct of the preference-based abstract argumentation framework and the redefined notion of a preferred extension and hence will not duplicate this work in its entirety. Interested readers are directed to [25] for the details. As the PAF extends from AF, the definition of AF is a natural point of departure. Hence, an abstract argumentation framework is defined as:

Definition 2.2.5 (Argumentation Framework). *An Abstract Argumentation Framework (AF) is a pair:*

$$AF = \langle AR, attacks_{AF} \rangle$$

where

- AR *is a set of arguments.*

- $attacks_{AF}$ *is a binary relations on AR, i.e. $attacks_{AF} \subseteq AR \times AR$.*

For readability, we will denote $attacks_{AF}(\alpha, \beta)$ to mean α attacks β. We also say that a set of arguments S attacks an argument β if β is attacked by an argument in S.

Definition 2.2.6 (Conflict-free, Defense, Admissible Semantics [25]). *Given $AF = \langle AR, attacks_{AF} \rangle$ and $A \subseteq AR$,*

- A *is conflict-free if and only if $\neg \exists \alpha, \beta \in A$ such that $attacks_{AF}(\alpha, \beta)$.*

- A *defends an argument $\alpha \in A$ if and only if $\forall \beta \in AR$, if $attacks_{AF}(\beta, \alpha)$ then $\exists \gamma \in A$ such that $attacks_{AF}(\gamma, \beta)$.*

- *A conflict-free set of arguments A is an admissible extension if and only if A defends all its elements.*

Given the definition of AF, conflict-freedom, defense and admissible semantics, a preference-based abstract argumentation framework is defined as:

Definition 2.2.7 (Preference-based Argumentation Framework [25]). *A Preference-based Abstract Argumentation Framework (PAF) is a triple:*

$$PAF = \langle AR, attacks_{PAF}, \geq \rangle$$

where

- *AR is a set of arguments.*

- *$attacks_{PAF}$ is a binary relation on AR, i.e. $attacks_{PAF} \subseteq AR \times AR$.*

- *$\geq$ is a (partial or total) reflexive, transitive (pre-order) on AR, i.e. $\geq \subseteq AR \times AR$.*

For readability, we will denote $attacks_{PAF}(\alpha, \beta)$ to mean α attacks β and $(\alpha, \beta) \in \geq$ or $\alpha \geq \beta$ to mean α is at least as strong as β.

Let us consider the motivating example 1.2.1

Example 2.2.5. *Assume that $\{\alpha, \beta, \gamma\}$ are arguments representing "Bill is a juvenile; therefore he should not go to jail", "Bill has assaulted someone, therefore he should be jailed" and "Bill has robbed someone, therefore he should be jailed" respectively. The $attacks_{PAF}$ relationship capturing the interplay between the arguments are represented as $\{attacks_{PAF}(\alpha, \beta),$ $attacks_{PAF}(\beta, \alpha),\ attacks_{PAF}(\alpha, \gamma)$ and $attacks_{PAF}(\gamma, \alpha)\}$. Let us assume that the argument "Bill is a juvenile; therefore he should not go to jail" is at least as strong as "Bill has robbed someone, therefore he should be jailed" and the "Bill has assaulted someone, therefore he should be jailed" is at least as strong as "Bill is a juvenile; therefore he should not go to jail". Hence the preference relationship between the arguments are represented as $\alpha \geq \gamma, \beta \geq \alpha$. Table 2.5 presents a summary of this discussion.*

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{PAF}$	$attacks_{PAF}(\alpha, \beta), attacks_{PAF}(\beta, \alpha)$ $attacks_{PAF}(\gamma, \alpha), attacks_{PAF}(\alpha, \gamma)$
$\geq$	$(\alpha, \gamma), (\beta, \alpha)$

This table provides a summary of the discussion in example 2.2.5 and illustrates the construction of PAF. AR is the set of arguments consisting of $\{\alpha, \beta, \gamma\}$. $attacks_{PAF}$ is the set of binary relation on AR capturing the relation between the arguments in AR. $\geq$ is the preference relation placed on the set of arguments.

Table 2.5: Summary of Example 2.2.5

Utilising a digraph, we can illustrate the interaction between the arguments by representing arguments as labelled vertices and the $attacks_{PAF}$ relation as directed edges. Hence

attacks$_{PAF}$(α, β) is represented with a directed edge from the vertex α to β (see figure 2.2).

The arguments are represented by the three vertices α, β and γ. The directed edges represents the relations $attacks_{PAF}(\alpha, \beta)$, $attacks_{PAF}(\beta, \alpha)$, $attacks_{PAF}(\alpha, \gamma)$ and $attacks_{PAF}(\gamma, \alpha)$. Hence, the illustration can be interpreted as the argument α is attacking arguments β and γ while both arguments β and γ are attacking the argument α.

Figure 2.2: Graphical Illustration of Example 2.2.5

Bourguet et al. associated PAF to AF by replacing the $attacks_{AF}$ relation with defeats relation. The defeats relation is constructed by utilising the $attacks_{PAF}$ relation and the preferences on the argument. Such a construction allows for the refinement of the $attacks_{PAF}$ relation such that it only retains successful attacks.

Definition 2.2.8 (Defeats [25]). *Let $PAF = \langle AR, attacks_{PAF}, \geq \rangle$ be a preference-based argumentation framework. The AF associated with PAF is the pair $\langle AR, defeats_{PAF} \rangle$ where $defeats_{PAF} \subseteq AR \times AR$ such that $(\alpha, \beta) \in defeats_{PAF}$ if and only if $(\alpha, \beta) \in attacks_{PAF}$ and $(\beta, \alpha) \notin >$ where $(\alpha, \beta) \in >$ if and only if $(\alpha, \beta) \in \geq$ and $(\beta, \alpha) \notin \geq$.*

Given the definition associating PAF to AF, let us consider the following example.

Example 2.2.6. *In example 2.2.5, we have provided the arguments, the $attacks_{PAF}$ relationship and the preference relationship. Utilising the preferences relation, we can construct the defeat relation. Hence the following assertions are in the set of defeat relation: $defeat(\beta, \alpha)$, $defeat(\alpha, \gamma)$. Utilising a digraph, we can illustrate the interaction between the arguments by representing arguments as labelled vertices and the $attacks_{PAF}$ relation as directed edges. Hence $attacks_{PAF}(\alpha, \beta)$ is represented with a directed edge from the vertex α to β (see figure 2.3).*

Given the definition associating AF to PAF, we can now construct a new abstract argumentation system AF'. AF' consists of the set of arguments AR and the relation $defeats_{PAF}$. Given the definition of conflict-free, we are now able to determine the interaction within a set
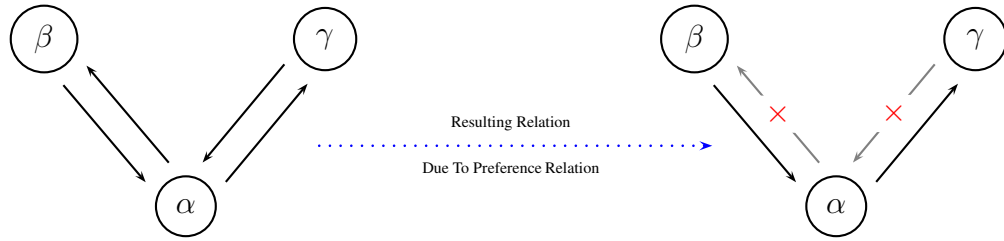
In figure 2.3, arguments are represented by the vertices α, β and γ. The directed edges represents the relations $attacks_{PAF}(\alpha, \beta)$, $attacks_{PAF}(\beta, \alpha)$, $attacks_{PAF}(\alpha, \gamma)$ and $attacks_{PAF}(\gamma, \alpha)$. $defeats_{PAF}$ relations represent successful attacks, therefore, unsuccessful attacks are eliminated. The eliminated directed edges are marked with a cross.

Figure 2.3: Graphical Illustration of Example 2.2.6

of arguments. A conflict-free *set of arguments is simply a set of arguments where arguments in the set can co-exists without defeating each other. Hence the following conflict-free set exists:* $\emptyset$, $\{\alpha\},\{\beta\},\{\gamma\}$ *and* $\{\beta, \gamma\}$ *for AF, PAF and AF′. An admissible set of arguments is conflict-free and defendable. Hence the following admissible sets exists:* $\emptyset$, $\{\alpha\},\{\beta\},\{\gamma\}$ *and* $\{\beta, \gamma\}$ *for AF,* $\emptyset$ *and* $\{\beta, \gamma\}$ *for PAF and AF′. Table 2.6 presents a summary of this discussion.*

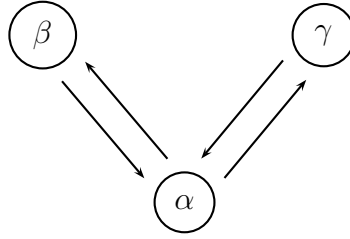|	AF	PAF	AF'
AR	$\{\alpha, \beta, \gamma\}$	$\{\alpha, \beta, \gamma\}$	$\{\alpha, \beta, \gamma\}$
$attacks_{PAF}$	$attacks_{PAF}(\alpha, \beta), attacks_{PAF}(\beta, \alpha)$ $attacks_{PAF}(\gamma, \alpha), attacks_{PAF}(\alpha, \gamma)$		-
$\geq$	-	$(\alpha, \gamma), (\beta, \alpha)$	-
$defeats_{PAF}$	-	$defeats_{PAF}(\beta, \alpha), defeats_{PAF}(\alpha, \gamma)$	
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$		
admissible	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$	$\emptyset, \{\beta, \gamma\}$	

Table 2.6: Summary of Example 2.2.6

Given the association between PAF to AF, Bourguet et al. extended the Value-based Abstract Argumentation Framework (VAF) proposed by Bench-Capon [22] and performed comparisons between PAF, AF and VAF. Again, we direct interested readers to [25] for the details.

2.2.3 Value-Based Abstract Argumentation Framework

Following from Perelman [108], Bench-Capon [22] argued that the use of argumentation in practical reasoning is to persuade rather than to prove, demonstrate or refute. In such

situations, persuasion relies on a recognition that the strength of an argument depends on the social values that it advances. Furthermore, Bench-Capon [22] highlighted that the success of an attack between arguments depends on the comparative strength of the values advanced by the arguments. Extending from the standard notion of Dung's AF, Bench-Capon [22] proposed the Value-based Abstract Argumentation Framework (VAF). Such a framework explores the acceptability properties of arguments, thus illustrated that disputants can concur on the acceptance of arguments, even when they differ on the importance of the values, hence, identifies points for which persuasion is possible.

In [22], Bench-Capon firstly introduced a global VAF for a set of audiences and followed by defining an audience specific value-based abstract argumentation framework. Bench-Capon [22] further introduced the notion of a set of audiences to individuated preferences ordering on the abstract values. Hence, there are potentially as many orderings on the set of abstract values as there are elements in the set of audiences. Our interest focuses on the preliminary structural definition of the audience specific VAF and the redefined notion of preferred extension and hence will not duplicate this work in its entirety. Interested readers are directed to [22, 49, 50, 78] for the details. Furthermore, without causing any confusion, we will dispense with the global VAF and refer directly to an audience specific VAF. As such, the VAF will be defined with respects to some audience. A Value-based Abstract Argumentation Framework is defined as:

Definition 2.2.9 (Value-based Argumentation Framework [22]). *A Value-based Abstract Argumentation Framework (VAF)is a 5-tuple:*

$$VAF = \langle AR, attacks_{VAF}, V, val, valpref \rangle$$

where

- *AR is a set of finite arguments.*

- *$attacks_{VAF}$ is a irreflexive binary relations on AR, i.e. $attacks_{VAF} \subseteq AR \times AR$.*

- *V is a nonempty set of abstract values.*

- *val is a function which maps elements of AR to elements of V*

- *$valpref$ is a transitive, irreflexive and asymmetric preference relation $valpref \subseteq V \times V$*

For convenience, given that $v_1, v_2 \in V$, we will denote $valpref(v_1, v_2)$ to mean v_1 is preferred to v_2.

In [52], arguments are assumed to have the same strength. This assumption is recognized to be too strong and often unsatisfied [25]. Recognising that the strength of an argument depends on the social values that it advances, Bench-Capon et al. [22,49,50,78] utilised abstract values to capture comparative strength of the values advanced by the arguments concerned. This extension provides a distinguish between success or failure of attacks. Hence, an argument α defeats another argument β if α attacks β and the value promoted by β is not more preferred than the value promoted by α. Bench-Capon defined the notion of defeats as:

Definition 2.2.10 (Defeats [22]). *Let $VAF = \langle AR, attacks_{VAF}, V, val, valpref \rangle$, an argument $\alpha \in AR$ defeats an argument $\beta \in AR$ ($defeats_{VAF}(\alpha, \beta)$) if and only if:*

$$attacks_{VAF}(\alpha, \beta) \wedge \neg valpref(val(\beta), val(\alpha)).$$

Let us consider the motivating example 1.2.1.

Example 2.2.7. *Assume that $\{\alpha, \beta, \gamma\}$ are arguments representing "Bill is a juvenile; therefore he should not go to jail", "Bill has assaulted someone, therefore he should be jailed" and "Bill has robbed someone, therefore he should be jailed" respectively. The $attacks_{VAF}$ relationship capturing the interplay between the arguments are represented as $\{attacks_{VAF}(\alpha, \beta),$ $attacks_{VAF}(\beta, \alpha), attacks_{VAF}(\alpha, \gamma)$ and $attacks_{VAF}(\gamma, \alpha)\}$. Assume also a set of abstract values $\{v_1, v_2\}$ with an ordering on the abstract values is represented as $\{valpref(v_1, v_2)\}$. Assume that Bill lives in a draconian society where crimes are punished. Hence, social values that the argument "Bill has robbed someone, therefore he should be jailed" promotes out weighs that of "Bill is a juvenile; therefore he should not go to jail". Similarly, the social values that the argument "Bill has assaulted someone, therefore he should be jailed" promotes out weighs that of "Bill is a juvenile; therefore he should not go to jail". Hence, the valuation that reflect such a view is: $val(\alpha) = v_2$, $val(\beta) = v_1$ and $val(\beta) = v_1$. Utilising a digraph, we can illustrate the interaction between the arguments by representing arguments as labelled vertices and the $attacks_{VAF}$ relation as directed edges. Hence $attacks_{VAF}(\alpha, \beta)$ is represented with a directed edge from the vertex α to β (see figure 2.4).*

Given the arguments and the $defeats_{VAF}$ relation, the conflict-free set of arguments are: $\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$. The resulting admissible set of arguments are: $\emptyset, \{\beta, \gamma\}$. Table 2.7 presents a summary of this discussion.

Similar to [10,25,77,78] , each VAF can be represented in an à la Dung style argumentation framework as $\langle AR, defeats_{VAF} \rangle$ where $defeats_{VAF} \subseteq AR \times AR$ such that $(\alpha, \beta) \in defeats_{VAF}$ if and only if $(\alpha, \beta) \in attacks_{VAF}$ and $\neg valpref(val(\beta), val(\alpha))$. Given the definition of
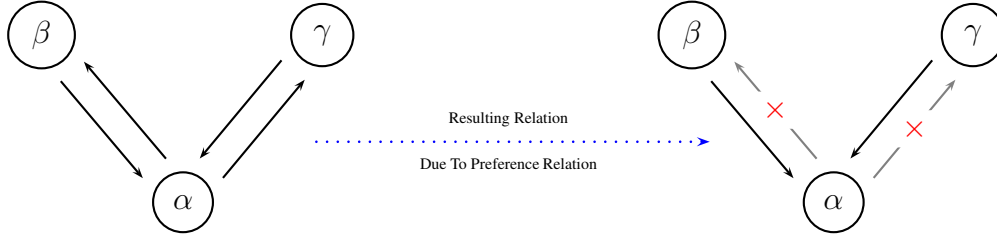
In figure 2.3, arguments are represented by the vertices α, β and γ. The directed edges represents the relations $attacks_{VAF}(\alpha, \beta)$, $attacks_{VAF}(\beta, \alpha)$, $attacks_{VAF}(\alpha, \gamma)$ and $attacks_{VAF}(\gamma, \alpha)$. $defeats_{VAF}$ relations represent successful attacks therefore, unsuccessful attacks are eliminated. The eliminated directed edges are marked with a cross.

Figure 2.4: Graphical Illustration of Example 2.2.7

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{VAF}$	$attacks_{VAF}(\alpha, \beta), attacks_{VAF}(\beta, \alpha)$ $attacks_{VAF}(\gamma, \alpha), attacks_{VAF}(\alpha, \gamma)$
V	$\{v_1, v_2\}$
val	$val(\alpha) = v_2, val(\beta) = v_1$ and $val(\beta) = v_1$
$valpref$	(v_1, v_2)
$defeats_{VAF}$	$\{defeats_{VAF}(\beta, \alpha), defeats_{VAF}(\gamma, \alpha)\}$

This table provides a summary of the discussion in example 2.2.7. AR is the set of arguments consisting of $\{\alpha, \beta, \gamma\}$. $attacks_{VAF}$ is the set of binary relation on AR capturing the $attacks_{VAF}$ relation between the arguments in AR. V is a set of abstract values.$valfpref$ is the preference relation placed on the set of arguments. The $defeats_{VAF}$ relation is then constructed based on the arguments, $attacks_{VAF}$ relation and preference relation.

Table 2.7: Summary of Example 2.2.5

$defeats_{VAF}$, Bench-Capon [22] continues the construction of VAF by defining acceptability, conflict-free and admissibility. The notion of acceptability is defined as:

Definition 2.2.11 (Acceptable [22]). *Given $VAF = \langle AR, attacks_{VAF}, V, val, valpref \rangle$ and a set of arguments $S \subseteq AR$. An argument $\alpha \in AR$ is* acceptable *to S ($acceptable(\alpha, S)$) if:*

$$(\forall \beta)((\beta \in AR \land defeats_{VAF}(\beta, \alpha)) \rightarrow (\exists \gamma)((\gamma \in S) \land defeats_{VAF}(\gamma, \beta))).$$

The notion of conflict-freedom is defined as:

Definition 2.2.12 (Conflict-free [22]). *Given $VAF = \langle AR, attacks_{VAF}, V, val, valpref \rangle$, a set of arguments $S \subseteq AR$ is* conflict-free *if:*

$$(\forall \alpha)(\forall \beta)((\alpha \in S \land \beta \in S) \rightarrow (\neg attacks_{VAF}(\alpha, \beta) \lor valpref(val(\beta), val(\alpha)))).$$

Given the notion of acceptability and conflict-freedom, admissibility is defined as:

Definition 2.2.13 (Admissible [22]). *Given $VAF = \langle AR, attacks_{VAF}, V, val, valpref \rangle$ and any given set of conflict-free arguments $S \subseteq AR$, S is* admissible *if:*

$$(\forall \alpha)((\alpha \in S) \rightarrow acceptable(\alpha, S)).$$

Given the notion of acceptability, let us consider the following example.

Example 2.2.8. *Continuing from example 2.2.7, we have provided the arguments, the $attacks_{VAF}$ relationship and the preference relationship. Utilising the preferences relation, we can construct the $defeats_{VAF}$ relation. Given the definition of conflict-free and acceptability, we are now able to determine the admissibility of a set of arguments. Hence the following admissible set exists: $\emptyset, \{\beta, \gamma\}$. Table 2.8 presents a summary of this discussion.*

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{VAF}$	$attacks_{VAF}(\alpha, \beta), attacks_{VAF}(\beta, \alpha)$ $attacks_{VAF}(\gamma, \alpha), attacks_{VAF}(\alpha, \gamma)$
V	$\{v_1, v_2\}$
val	$val(\alpha) = v_2, val(\beta) = v_1$ and $val(\beta) = v_1$
$valpref$	(v_1, v_2)
$defeats_{VAF}$	$\{defeats_{VAF}(\beta, \alpha), defeats_{VAF}(\gamma, \alpha)\}$
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
admissible	$\emptyset, \{\beta, \gamma\}$

This table provides a summary of the discussion in example 2.2.8. AR is the set of arguments consisting of $\{\alpha, \beta, \gamma\}$. $attacks_{VAF}$ is the set of binary relation on AR capturing the relation between the arguments in AR. V is a set of abstract values.$valfpref$ is the preference relation placed on the set of arguments. The $defeats_{VAF}$ relation is then constructed based on the arguments, $attacks_{VAF}$ relation and preference relation. The conflict-free sets of arguments and admissible sets of arguments can then be determined utilising the set of arguments and the $defeats_{VAF}$ relation.

Table 2.8: Summary of Example 2.2.8

Given the notion of admissibility, a preferred extension is defined as:

Definition 2.2.14 (Preferred Extension [22]). *Given $VAF = \langle AR, attacks_{VAF}, V, val, valpref \rangle$ and any given set of arguments $S \subseteq AR$, S is a* preferred extension *if it is a maximal (wrt set inclusion) admissible S.*

Given the notion of a preferred extension, let us consider the following example.

Example 2.2.9. *Continuing from example 2.2.8, we have provided the arguments, the $attacks_{VAF}$ relationship and the preference relationship. Utilising the preferences relation, we can construct the defeat relation. Utilising the defeat relation, the conflict-free and admissible sets of arguments can be determined. Given the definition of a preferred extension, we are now able to determine the preferred set of arguments. Hence, $\{\beta, \gamma\}$ is the only preferred extension exists. Table 2.9 presents a summary of this discussion.*

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{VAF}$	$attacks_{VAF}(\alpha, \beta), attacks_{VAF}(\beta, \alpha)$ $attacks_{VAF}(\gamma, \alpha), attacks_{VAF}(\alpha, \gamma)$
V	$\{v_1, v_2\}$
val	$val(\alpha) = v_2, val(\beta) = v_1$ and $val(\beta) = v_1$
$valpref$	(v_1, v_2)
$defeats_{VAF}$	$\{defeats_{VAF}(\beta, \alpha), defeats_{VAF}(\gamma, \alpha)\}$
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
admissible	$\emptyset, \{\beta, \gamma\}$
preferred	$\{\beta, \gamma\}$

This table provides a summary of the discussion in example 2.2.9. AR is the set of arguments consisting of $\{\alpha, \beta, \gamma\}$. $attacks_{VAF}$ is the set of binary relation on AR capturing the relation between the arguments in AR. V is a set of abstract values.$valfpref$ is the preference relation placed on the set of arguments. The $defeats_{VAF}$ relation is then constructed based on the arguments, $attacks_{VAF}$ relation and preference relation. The conflict-free sets of arguments and admissible sets of arguments can then be determined utilising the set of arguments and the $defeats_{VAF}$ relation. The preferred extension is selected from the set of admissible sets of argument.

Table 2.9: Summary of Example 2.2.9

Given the definitions for VAF and a preferred extension, Bench-Capon performs a discussion on acceptability within VAF. Two notions of acceptabilites are introduced: objective acceptability and subjective acceptability. Objective acceptability describes acceptance arguments by all audience participating in the debate and subjective acceptability describes acceptance of arguments by some audience. Note that these notions differ to that introduced in 1.2.2.

2.3 Theory Change

In logic, logically closed sets are called "theories" and in formal epistemology, they are referred to as "knowledge sets", or commonly "belief sets". Hence in this discussion, we will use these terms interchangeably. The topic of theory change is a subject of much discussion in the knowledge representation and reasoning community [2, 3, 61–63, 67, 82, 83, 132–134].

This topic of theory change emerged as a result of the convergence of the research in traditional philosophy and the development of Artificial Intelligence in computing science.

The capstone result is the AGM model proposed by Carlos Alchourrón, Peter Gärdenfors, and David Makinson [3]. The AGM model provides a general and versatile formal framework for studies of belief change. Following the proposal by Alchourrón et al., the major concepts and constructions of the AGM model have been subjected to significant elaboration and development, forming the core of current belief revision theory. In the AGM model (as well as in most other models) of belief change, beliefs are represented by sentences in some formal language. As is usual in logic, the formal language is identified with the set of assertions and the usual connectives: negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), implication ($\rightarrow$), and equivalence ($\leftrightarrow$). $\perp$ denotes an arbitrary contradiction and $\top$ an arbitrary tautology. Furthermore, given any set of assertions A, $Cn(A)$ is the set of logical consequence of A.

The AGM framework prescribes postulates on three types of belief change operators: expansion, contraction and revision. A belief set is generally denoted as K and $+$, $\div$ and $*$ denotes the expansion operator, contraction operator and revision operator respectively.

2.3.1 AGM Expansion

In expansion, the sentence (p) is added to a belief set (K) without retracting any existing beliefs such that the sentence (p) is a consequence of the belief set. In other words, $K + p = Cn(K \cup \{p\})$. Expansion is recommended only if p is consistent with K. The basic AGM postulates for expansion are:

1. (Closure) $K + p = Cn(K + p)$

2. (Success) $p \in Cn(K + p)$

3. (Inclusion) $K \subseteq K + p$

4. (Vacuity) If $p \in Cn(K)$, then $K + p = K$

5. (monotonicity) if $K \subseteq H$, then $K + p \subseteq H + p$

6. (minimality) For all belief sets K and all sentence p, $K + p$ is the smallest belief set that satisfies postulates 1–5

The postulate of closure expresses the fact that the expansion operator takes a belief set and a sentence as input and produces a belief set. Essentially, the closure postulate expresses the principle of category matching. The success postulate states that the input sentence is accepted in the expansion and is a consequence of the expanded belief set. The inclusion postulate says that no sentence in the belief set is retracted when an expansion is performed. This postulate expresses a form of the principle of minimal change. The vacuity postulate represents a boundary case and states that nothing needs to be done if the input sentence is already accepted. The postulate of monotonicity expresses that if one belief set contains at least the same information as another, then its expansion will contain at least the information of the expansion of the other with respect to the same input sentence. The postulate of minimality states that the smallest possible change to accommodate the new information is made. The term "smallest" is understood with respect to set inclusion (w.r.t the original belief set). This postulate can also be viewed as an expression of the principle of minimal change

2.3.2 AGM Contraction

In contraction, a specified sentence (p) is removed from the set of beliefs (K) such that the sentence is not a consequence of the belief set. Furthermore, the outcome of the contraction operation should not unnecessarily remove elements of K. Hence, $K \div p$ should produce the maximal subset of K that does not entail p. The basic AGM postulates for contraction are:

1. (Closure) $K \div p = Cn(K \div p)$

2. (Success) If p is not an arbitrary tautology, then $p \notin Cn(K \div p)$

3. (Inclusion) $K \div p \subseteq K$

4. (Vacuity) If $p \notin Cn(K)$, then $K \div p = K$

5. (extensionality) If $p \leftrightarrow q$, then $K \div p = K \div q$

6. (Recovery) $K \subseteq (K \div p) + p$

7. (Conjunctive Overlap) $(K \div p) \cap (K \div q) \subseteq K \div (p \wedge q)$

8. (Conjunctive Inclusion) If $p \notin K \div (p \wedge q), then K \div (p \wedge q) \subseteq K \div p$

Similar to the postulate of closure for belief set expansion, the postulate of closure expresses that the outcome of the contraction operation should be logically closed. The success postulate states that the contraction of a belief set K by p should result in a belief set that does not entail p. However, this condition is too restrictive as it will also exclude logical truth. Hence the success postulate is conditional on p not being a tautology. The inclusion postulate says that no contracted belief set is a subset of the original belief set and is a form of the principle of minimal change. The vacuity postulate represents a boundary case and states that nothing needs to be done if the belief set does not contain the sentence to be contracted. The postulate of extensionality allows for logically equivalent sentences to be freely substituted. This postulate expresses the principle of irrelevance of syntax. The recovery postulates states that so much of the belief set is retained after the contraction of p such that everything can be recovered by the re-inclusion of p. This postulate can be considered as an expression of the principle of minimal change as this principle dictates that beliefs are not to be unnecessarily given up. Recovery is arguably the most controversial of the AGM rationality postulates and there are a number of contributions discussing its removal [58, 93]. The remaining two postulates are supplementary postulates. The postulate of conjunctive overlap states that, the result of contraction by $p \wedge q$ should consists the result of contraction by p and contraction by q. The postulate of conjunctive inclusion states that to give up the sentence $p \wedge q$, it is required to either contract p or contract q or both from the belief set.

2.3.3 AGM Revision

In revision, a sentence (p) is added to the belief set (K). Should an inconsistency exists due to the addition of p, other sentences are removed from the set of beliefs, such that the p is still a consequence of the belief set. Hence the revision operator performs two main tasks: to add a new belief to the belief set and to ensure the resulting belief set is consistent. The AGM postulates for revision are:

1. (Closure) $K * P = Cn(K * p)$

2. (Success) $p \in K * p$

3. (Inclusion) $K * p \subseteq K + p$

4. (Vacuity) If $(\neg p) \notin K$, then $K * p = K + p$

5. (Consistency) $K * p$ is inconsistent only if p is inconsistent

6. (extensionality) If $p \leftrightarrow q$, then $K * p = K * q$

7. (Superexpansion) $K * (p \wedge q) \subseteq (K * p) + q$

8. (Subexpansion) If $(\neg q) \notin K * p$ then $(K * p) + q \subseteq K * (p \wedge q)$

Similar to the postulate of closure for belief set contraction, the postulate of closure for belief set revision expresses that the outcome of the contraction operation should be logically closed. The success postulate states that the revision of a belief set K by p should result in a belief set that entails p. The inclusion postulate says that revising the belief set K by p is a subset of the belief set K expanded by p. The vacuity postulate represents a boundary case and states that only expansion needs to be performed should the result from the expansion be consistent. In other words, only an expansion needs to be performed if the belief set does not contain the negation of the sentence to be revised. The postulate of consistency states that an inconsistency in a belief set can only be introduced by the revising sentence as the initial belief set is internally consistent. Similar to the postulate of extensionality for belief set contraction, the postulate of irrelevance of syntax for belief set revision allows for logically equivalent sentences to be freely substituted. Two supplementary postulates for revision exist. The superexpansion postulate states that any belief included in the revision $K * (p \wedge q)$ should also be included if we first revise by p and then expand the result by q. The subexpansion postulate says that, if $\neg q$ is not in the result of $K * p$ then any belief included by first revising K by p and expanding the result by q should also be included in the revision of K by $(p \wedge q)$.

2.4 Mixed-Initiative Interaction

The topic of mixed-initiative interaction is a subject of much discussion in the artificial intelligence research community [4, 45, 46, 48, 66, 73, 74, 102, 141, 144]. Mixed-initiative artificial intelligence systems have been designed for a variety of applications areas, including robotics [59], planning [27, 30, 57, 130, 131] and tutoring [32].

As highlighted by Horvitz [74], the area of research on mixed-initiative interaction is relatively nascent. Hence, researchers within the field of mixed-initiative systems have yet to arrive at a consensus about what constitutes *initiative*. Debate exists on differing opinions about whether the initiative is taken simply to direct the dialogue or to alter the course of the problem solving [43, 44]. However, the general focus of mixed-initiative interaction research centres around the development of methods that enable computing systems to support an efficient, natural interleaving of contributions between participating parties with the aim of

converging onto solutions to problems. Cohen et al. [44] highlighted that mixed-initiative systems involved an ongoing dialogue between the user and the system, so that initiative is taken as part of deliberately interacting with the other party.

Mixed-initiative interaction can be summarised as a flexible interaction strategy in which parties (human or computer) take initiatives in the contribution of their best suited-skill set at the most appropriate time to solving a problem, achieving a goal or coming to a joint understanding [4, 44, 74]. In other words, such a system would carry out problem solving tasks on behalf of user where both the user and the system can take the initiative by directing the problem solving task. Noted by Allen [4], mixed-initiative systems need not involve a human.

Early studies into the design of mixed-initiative systems focus around dialogue systems [152], with the aim to model mixed-initiative discourse using shifting in control associated with linguistic construction. Hence, the model proposed in [152] assists in the analysis of interactions in the discourse. Chu-Carroll et al. [42] make distinctions between task initiative from dialogue initiative and present an evidential model for tracking shifts in both types of initiatives in collaborative dialogue interactions. Utterances are classified according to the following types: assertion, command, question and prompt. Such classification naturally leads into studies to determine the appropriate situations to interrupt a dialogue. Thus, the ability to recognise problem-solving opportunities is outside the current scope or focus of attention.

Considerations in the use of mixed-initiative interaction have also been proposed in planning [1,26,27,30,31,54–57]. Studies such as TRAINS-95 [57,137], led to the conclusion that the plan reasoning requirements in mixed-initiative systems differs from that of traditional planning and highlighted that an interactive, dialogue-based approach to plan reasoning is effective. As mixed-initiative reasoning is concerned with the development of collaborative system in which the human and automated agents work together to achieve a common goal in a way that exploits their complementary capabilities, several qualities for a mixed-initiative systems personalisability, directability, teachability and transparency of operation are desirable [141]. Furthermore, the development of mixed-initiative system is challenging due to the fact that it requires the synergistic integration of many areas of artificial intelligence ranging from knowledge representation and reasoning through to human-computer interaction. Tecuci et al. [141] considered seven aspects or issues of mixed-initiative reasoning to guide general design principles and methods:

1. Task issues.

2. Control issues.

3. Awareness issues.

4. Communication issues.

5. Personalisation issues.

6. Architecture issues.

7. Evaluation issues.

The task issues covers the division of responsibility between the human and the agent for the task that needs to be performed. One dimension of such complementarity division may relate to their individual strength and weaknesses. For example, the difference between a human's reasoning styles and the agent's computational strengths could allow humans to perform problem solving and decision making while the computational strength of automated agents allows them to perform complex mathematical computation, storing and retrieval of large quantity of data.

The control issues relates to the strategies for shifting the initiative and control between the human and the agent in a proactive and reactive manner. Decisions on the division of labour not only depends on the qualifications of the participants but also on the set of tasks that need to be performed at the appropriate time. Horvitz [73] highlighted several deficiencies and proposed a set of design principles impacting on control issues.

The awareness issues relates to the maintenance of a shared understanding between the human and the agent on the constantly evolving state of the problem-solving process. For collaboration between human and agent to be achieved, the collaborating parties need to share facts and beliefs. Furthermore, there should be a common understanding of their joint goals. Tecuci et al. [141] illustrated that this is an issue due to the fact that humans and automated agents differ in their interaction modalities and comprehension capabilities.

The communication issues relates to the protocols that facilitate the exchange of knowledge and information between the human and the agent. Such protocol may include mixed-initiative dialogue and multi-modal interfaces [141]. When choosing the form of communicating between human and agents, several consideration should be made. Firstly, consideration should be made on the efficiency of the means of communication. Secondly, consideration should be made on the ease of use. Finally, the contribution to the level of mixed-initiative collaboration.

The personalisation issues relates to the adaptation of agent's knowledge and behaviour to its user's problem solving strategies, preferences, biases and assumptions. The consideration of a user's preferences and biases is particularly important as it enable systems to produce solutions that are likely to be acceptable or desirable to the user and helps the user to avoid mistakes by checking for biases and assumptions.

The architecture issues relates to the design principles, methodologies, and technologies for different types of mixed-initiative roles and behaviours. Tecuci et al. [141] suggested that by identifying and studying different types of mixed-initiative roles and behaviours significantly facilitate the development of useful mixed-initiative systems and acceptance of such systems.

The evaluation issues are related to the human and automated agent contribution to the emergent behaviour of the system and the overall system's performance versus fully automated, fully manual, or alternative mixed-initiative approaches. As highlighted by Tecuci et al. [141] and Kirkpatrick [84], mixed-initiative systems are very difficult to evaluate. This is due to several reasons. Firstly, mixed-initiative systems are generally very complex consisting of components for reasoning, communication, planning and learning. The complex intertwining of components subsequently makes the system very difficult to evaluate. Secondly, the evaluation process may require different types of users. These resources may not be immediately available. Thus, further complicating the sequencing of these intertwined components. Hence, resulting in a costly and time consuming evaluating process. Finally, the evaluation requires several comparisons between fully automated, fully manual, or alternative mixed-initiative approaches.

Given the design consideration, the degree of interaction can be decomposed into several different levels. Allen [4] presented four levels of mixed-initiative interaction:

- Unsolicited reporting.

- Sub-dialogue initiation.

- Fixed subtask mixed-initiative.

- Negotiated mixed initiative.

The unsolicited reporting level of interaction describes the ability of an agent to notify others of critical information as it arises. This level of interaction is considered the first step toward mixed-initiative where by the agent notifies the user of a change in situation, plan or identifies a problem. At this level of mixed-initiative interaction, the agent does not coordinate the subsequent interaction.

The sub-dialogue initiation level of interaction describes the ability for an agent to initiate sub-dialogues to perform clarification and corrects. At this level of interaction, under certain situation, the agent might initiate a sub-dialogue such as asking for clarification. Such clarification might take several interactions and hence during these interactions, the agent has temporarily taken the initiative until the issue is clarified.

The fixed subtask mixed-initiative level of interaction describes the ability for an agent to take initiatives in solving or performing predefined subtasks. At this level of interaction, the agent has the responsibility to perform certain sub-tasks. While the agent is working on these operations, it is maintaining the initiative. Once the sub-task is completed, the initiative is reverted back to the user.

The negotiated mixed initiative level of interaction describes the ability for an agent to co-ordinate and negotiate with other agents to determine initiative or division of labour for achieving goals and tasks. At this level of interaction, there is no fixed assignment of responsibilities or initiatives. Each agent constantly monitors the current task and evaluates whether it should take the initiative in the interaction.

In following section, we will present a summary of the main concepts this discussed in this chapter.

2.5 Summary

In this chapter, we have presented an overview on the current state-of-the-art in formal argumentation theory, belief theory change and mixed-initiative interaction. We have also presented the three abstract argumentation frameworks (AF, PAF, VAF) in which our proposed framework utilised as the foundational preliminaries. The next two chapter will introduce the formal aspect of this dissertation. In chapter 3, we will present the abstract preference-based accrual argumentation framework ($PAAF$). The aim of preferences-based accrual abstract argumentation framework is to formalise the use of preferences in an abstract argumentation framework to address issues presented in chapter 1. In chapter 4, we will present the mixed-initiative argumentation framework ($MIAF$). The mixed-initiative argumentation framework performs argumentation theory change by considering revisions on the conflict and preference relations. This capability coupled with mixed-initiative interaction between users and systems allows for the management of justifications over a sequence of decisions.

Part II

Abstract Argumentation Frameworks & Applications

3

Preference-Based Accrual Argumentation Framework

"The moment we want to believe something,
we suddenly see all the arguments for it,
and become blind to the arguments against it."
– George B. Shaw

3.1   Introduction

In chapter 2, we introduced the abstract argumentation framework (AF) proposed by Dung in [52], the preference-based argumentation framework (PAF) proposed by Bourguet et al. [25] and the value-based argumentation framework (VAF) proposed by Bench-Capon in [22] as the foundations of our work. In this chapter, we will introduce a novel abstract preferences-based accrual argumentation framework ($PAAF$).

There have been considerable debates [65, 87, 117, 123, 124, 145–147, 149, 151] about when it is appropriate to perform accrual of arguments and if a general theory for the accrual of arguments can be formulated. We feel that the choice of whether or not to perform accrual is domain-specific and is dependent on an individual's acceptability criteria or preferences. The main focus of this chapter is not to address this debate, but to define an abstract machinery for accrual of arguments. We show how accrual of arguments can be preformed in an abstract framework and the influence of individual preferences on the acceptability of such arguments. When argumentation is used as a conflict resolution technique in multi-agent settings, agents are required to be able to determine the "believability" of arguments given to them based on their own knowledge base. Associated with these knowledge bases are preferences specification unique to each agent. In group decision support applications using argumentation, the ranking of sets of arguments becomes a critical component of the decision making machinery as arguments provide the basis for supporting or attacking of decisions. Hence, the ranking permits the arguments to play a more significant role as well as informing the reasoner as to which is the better choice.

In section 3.1.1, we will motivate the need for an argumentation system to address issues of argument strength when accrual is of interest. In section 3.2, we will introduce accrual in an abstract argumentation framework by extending the abstract argumentation framework (AF), preference-based argumentation framework (PAF) and the value-based argumentation framework (VAF) to construct an abstract preferences-based accrual argumentation framework ($PAAF$). Utilising a running example, we will illustrate the ability of our framework to perform accrual of arguments and highlight some of the unique features of this framework. Section 3.3 presents several key concepts. Firstly, a brief discussion and comparison illustrating the differences between the AF, VAF and $PAAF$ is presented. Secondly, we present a discussion addressing issues related to argument source, ownership, credibility and trust as illustrated in the motivating examples. In section 3.4, we present a summary of key ideas and concepts presented in this chapter.

3.1.1 Illustrating Example

In section 1.2.2, we demonstrated that typically, the accrual of arguments strengthens an argumentation position. In this example, we will show that accrual of arguments can in fact also weaken an argumentation position. Consider an adaptation of the example from [122, 123] giving two reasons against jogging. For a particular runner (Tom), the combination of heat and rain may be less unpleasant than heat or rain alone. In such a situation, the accrual

of arguments weakens the conclusive strength of a set of arguments. Let us assume that Tom has the goal to be fit. Two independent argument exchanges are presented as follows:

> A_1:*Tom has not been jogging for several days, so he should go jogging.*
> A_2:*It is raining, so Tom should not go jogging.*

From the previous description, it is clear that Tom believes that it is unpleasant to run in the rain. It might be intuitive to conclude that the second argument A_2 is more preferred. Now consider the following exchange:

> A_1:*Tom has not been jogging for several days, so he should go jogging.*
> A_3:*It is hot, so Tom should not go jogging.*

Again, it might be intuitive to conclude that the second argument A_3 is more preferred. Hence A_3 is more preferred. However, if we were to consider the two exchanges simultaneously, the conclusion will be different.

> A_1:*Tom has not been jogging for several days, so he should go jogging.*
> A_2:*It is raining, so Tom should not go jogging.*
> A_3:*It is hot, so Tom should not go jogging.*

From the initial description, we are aware that the presence of the rain and heat together weakens the argument against jogging. Thus, the first argument (A_1) defeats the accrual of the two latter arguments (A_2,A_3). This example also highlights the influence preferences have on the acceptability of arguments. Note that the accrual of arguments is a meta-level operation and is dependent on the semantics of the arguments rather than the structure (syntax) of the arguments.

In this spirit, we will discuss in this chapter how an abstract preference-based accrual argumentation framework can be used to model such disagreements and provide techniques for persuasion in such a context.

3.2 Formal Framework

In this section, we will present the formal details for the abstract preference-based accrual argumentation framework ($PAAF$). Formally, an abstract preference-based accrual argumentation framework is defined as:

Definition 3.2.1. *An abstract preference-based accrual argumentation framework ($PAAF$) is a triple:*

$$PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$$

where

- AR *is a set of arguments.*

- $attacks_{PAAF}$ *is a binary relation on AR (i.e., $attacks_{PAAF} \subseteq AR \times AR$).*

- $Bel = \langle V, \leq, \Phi \rangle$ *where*

    - V *is a set of abstract values.*

    - $\leq$ *is a total ordering on V.*

    - Φ *is a total valuation function which maps elements of 2^{AR} to elements of V (ie. $\Phi : 2^{AR} \longrightarrow V$).*

For readability, we will use $attacks_{PAAF}(\alpha, \beta)$ to denote α attacks β. Similarly, given $v_1, v_2 \in V$, we will use $pref(v_1, v_2)$ to denote $v_1 \leq v_2$ (i.e., v_1 is preferred to v_2). Note that without lost of generality, for the rest of this chapter, we will use the terms abstract values and preference values interchangeable.

Drawing from the belief revision literature, the symbol Bel is traditionally used to denote a belief set. In $PAAF$, Bel represents a three-tuple consisting of preference values, an ordering on the preference values and the valuation function that assigns preference values to arguments. This three-tuple represents an agent's or individual's preferences on a given set of arguments. These preferences reflect their commitment to the arguments. Hence, is synonymous to the traditional notion of a belief set where a belief set consisting of sentences that an agent is committed to believing as highlighted by Isaac Levi [88].

Although there exist great structural similarities between $PAAF$ and the abstract argumentation framework (AF), the preference-based argumentation framework (PAF) as well as the

value-based argumentation framework (VAF), neither AF, PAF nor VAF provide an explicit account of accrual of arguments and the associated reasoning with preferences. There are two substantial differences which allows this framework to explicitly capture the accrual in abstract argumentation.

Firstly, this approach differs from that proposed by Bench-Capon [22], Kaci et al. [77, 78] and Bourguet et al. [25]. In VAF and PAF, "abstract values" have the highest priority. As such, "abstract values" are utilised to overrule arguments at all times hence inducing an acceptability criteria reflecting the ordering on the "abstract values". Our aim is not to overrule the underlying argumentation machinery but to provide additional reasoning capabilities when a tie is encountered. One could view this addition as an additional "plug-in" to any exiting argumentation system. This is achieved by retaining the essence of the underlying attacks relation and refining the attacks relation into a definition of the defeats relation. This defeats relation differs substantially from that presented by Bench-Capon, Kaci et al. and Bourguet et al.. Secondly, we extend the framework by redefining the valuation function. This modification provides $PAAF$ with an addition to the available semantics, hence explicitly capturing the notion of accrual as well as ordering argumentation outcomes based on preferences. To achieve this, we define the valuation function that maps between the set of abstract values and the power set of the arguments in place of the original function as defined by Bench-Capon in [22]. This valuation function is the inverse of that defined by Bourguet et al. in [25]. Let us illustrate the use of $PAAF$ by translating the example 3.1.1 into the framework.

Example 3.2.1. *Assume that $\{\alpha, \beta, \gamma\}$ are arguments represents $\{$"Tom has not been jogging for several days, so he should go jogging.", "It is raining, so Tom should not go jogging." and "It is hot, so Tom should not go jogging."$\}$ respectively. The $attacks_{PAAF}$ relationship capturing the interplay between the arguments is represented as $\{attacks_{PAAF}(\alpha, \beta),$ $attacks_{PAAF}(\beta, \alpha), attacks_{PAAF}(\alpha, \gamma), attacks_{PAAF}(\gamma, \alpha)\}$. Table 3.1 presents a summary of this discussion.*

	Exchange 1	Exchange 2	Exchange 3
AR	$\{\alpha, \beta\}$	$\{\alpha, \gamma\}$	$\{\alpha, \beta, \gamma\}$
$attacks_{PAAF}$	$attacks_{PAAF}(\alpha, \beta)$ $attacks_{PAAF}(\beta, \alpha)$	$attacks_{PAAF}(\gamma, \alpha)$ $attacks_{PAAF}(\alpha, \gamma)$	$attacks_{PAAF}(\alpha, \beta), attacks_{PAAF}(\beta, \alpha)$ $attacks_{PAAF}(\gamma, \alpha), attacks_{PAAF}(\alpha, \gamma)$

This table provides a summary of the discussion in example 3.2.1 and illustrates how arguments and their attack relations are represented in $PAAF$.

Table 3.1: Summary of Example 3.2.1

Futhermore, utilising a digraph, we can illustrate the interaction between the arguments by representing arguments as labelled vertices and the attack relation as directed edges. Hence

$attacks_{PAAF}(\alpha, \beta)$ *is represented with an directed edge from the vertex α to β. Each exchanges of arguments can be captured as separate digraphs (see figure 3.1).*

(a) Exchange 1   (b) Exchange 2   (c) Exchange 3
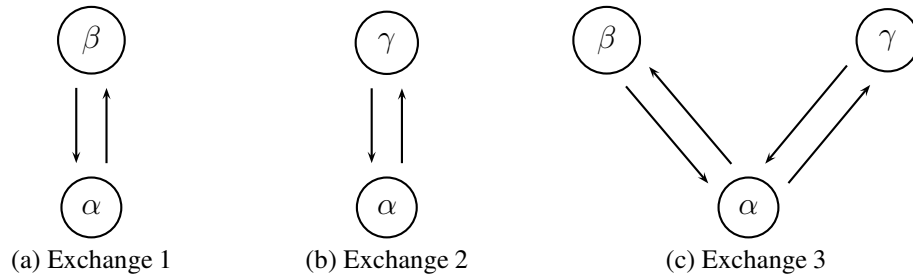
In figure 3.1a, figure 3.1b and figure 3.1c, arguments are represented by the vertices α, β and γ. The directed edges represent the relations $attacks_{PAAF}(\alpha, \beta)$ and $attacks_{PAAF}(\beta, \alpha)$, $attacks_{PAAF}(\alpha, \gamma)$ and $attacks_{PAAF}(\gamma, \alpha)$, and $attacks_{PAAF}(\alpha, \beta)$, $attacks_{PAAF}(\beta, \alpha)$, $attacks_{PAAF}(\alpha, \gamma)$ and $attacks_{PAAF}(\gamma, \alpha)$ respectively. Hence, figure 3.1a is interpreted as: α is attacking β and β is attacking α. Figure 3.1b is interpreted as: α is attacking γ and γ is attacking α. Figure 3.1c is interpreted as: α is attacking β and γ while both β and γ are attacking α separately.

Figure 3.1: Graphical Illustration of Example 3.1.1

Given the abstract framework, we will now provide a notion of a *conflict-free* set of arguments. A *conflict-free* set of arguments is simply a set of arguments where arguments in the set do not attack each other:

Definition 3.2.2. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$, a set of arguments S is said to be* **conflict-free** *if and only if:*

$$\neg(\exists\alpha\exists\beta((\alpha \in S) \wedge (\beta \in S) \wedge attacks_{PAAF}(\alpha, \beta))).$$

Given the definition of conflict-free, let us consider the following example.

Example 3.2.2. *Continuing from example 3.2.1, the following conflict-free set exists: The empty-set and $\{\alpha\}$ exists in all three exchanges. $\{\beta\}$ exists in exchange 1 and 3. The set $\{\gamma\}$ exists in exchange 2 and 3. Last but not least, the set $\{\beta, \gamma\}$ exists in exchange 3. Table 3.2 presents a summary of this discussion.*

In the following definition of defeat, we will utilise the two functions to extract from a set of arguments the subset that is *relevant* in relation to another set of arguments, in particular, the subset of arguments that participates in an attack and the subset of arguments that are under attack. We need to focus attention on the *relevant subsets* because our approach evaluates

	Exchange 1	Exchange 2	Exchange 3
AR	$\{\alpha, \beta\}$	$\{\alpha, \gamma\}$	$\{\alpha, \beta, \gamma\}$
$attacks_{PAAF}$	$attacks_{PAAF}(\alpha, \beta)$ $attacks_{PAAF}(\beta, \alpha)$	$attacks_{PAAF}(\gamma, \alpha)$ $attacks_{PAAF}(\alpha, \gamma)$	$attacks_{PAAF}(\alpha, \beta), attacks_{PAAF}(\beta, \alpha)$ $attacks_{PAAF}(\gamma, \alpha), attacks_{PAAF}(\alpha, \gamma)$
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}$	$\emptyset, \{\alpha\}, \{\gamma\}$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$

This table provides a summary of the discussion in example 3.2.2 and illustrates the computation of conflict-free sets of arguments.

Table 3.2: Summary of Example 3.2.2

sets of arguments with preference values, and without this restriction, the preference value of a given set of arguments could be artificially inflated by including highly preferred but irrelevant arguments. Thus, we define Θ with subscripts a and u representing *attacking* and *under-attacked* arguments respectively. Hence, given two sets of arguments $A, B \subseteq AR$, $\Theta_a(A, B)$ returns a subset of arguments from A that attacks some argument in B and $\Theta_u(A, B)$ returns a subset of arguments from A that is under attack from some argument in B.

Definition 3.2.3. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and $A, B \subseteq AR$ be two sets of arguments. Then,*

$$\Theta_a(A, B) = \{\alpha | \alpha \in A \wedge \exists \beta \in B \text{ s.t. } attacks_{PAAF}(\alpha, \beta)\}.$$

Definition 3.2.4. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and $A, B \subseteq AR$ be two sets of arguments. Then,*

$$\Theta_u(A, B) = \{\alpha | \alpha \in A \wedge \exists \beta \in B \text{ s.t. } attacks_{PAAF}(\beta, \alpha)\}.$$

A key motivation for our definition of defeat is to preserve the original interplay between arguments and only utilise preferences to intervene when no clear winners are obvious. Intuitively, our notion of defeat consists of three basic components. We will define three binary relations *defensible*, *overpower* and *preferable* between sets of arguments. Given two sets of arguments A and B, the notion of defensible simply means that every attacking argument from B is counter-attacked by some argument in A.

Definition 3.2.5. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and $A, B \subseteq AR$ be two sets of arguments. A is **defensible** against B (denoted as **defensible**(A, B)) if and only if:*

$$\Theta_a(B, A) \subseteq \Theta_u(B, A).$$

Given two sets of arguments A and B, the notion of A overpowering B simply means that B is attacked by some arguments in A and there exist some arguments in A that are not counter-attacked by arguments in B.

Definition 3.2.6. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and $A, B \subseteq AR$ be two sets of arguments. A **overpowers** B (denoted as **overpower**(A, B)) if and only if:*

$$\Theta_a(A, B) \setminus \Theta_u(A, B) \neq \emptyset.$$

One of the principle of accrual highlighted by Prakken [122] states that the accrual of a set of arguments may sometimes be weaker than the arguments in the set. In $PAAF$, the use of preferences as part of the definition of defeat and the notion of maximal defeaters allows for such a property. Furthermore, utilising preferences in such a manner can capture the notion of undercutting as highlighted by Pollock [109], Prakken and Vreeswijk [125] where there exists a choice between accepting or denying the relation between premises and conclusions in a non-deductive argument. Note that the preference ordering can be informed by some external source and hence can provide the rationale for a choice.

Given two abstract values v_1 and v_2, we can construct three possible preference comparison outcomes: v_1 is more preferred to v_2; v_2 is more preferred to v_1; v_1 and v_2 are indifferent (equally preferred). Since the use of preference in our framework is to break ties, our aim is to provide a relaxed definition of "preferred" such that it captures situations where a value is more preferred as well as when the two values are indifferent. So, we can express the desired condition as $pref(v_1, v_2) \vee (pref(v_1, v_2) \wedge pref(v_2, v_1)) \equiv pref(v_1, v_2)$

Definition 3.2.7. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and $A, B \subseteq AR$ be two sets of arguments. A is **preferable** to B (denoted as **preferable**(A, B)) if and only if:*

$$pref(\Phi(A), \Phi(B)).$$

Using the notion of defensible and overpowers, we are able to determine defeat if there exists an obvious winner. Using the notion of defensible and preferable, we are able to determine defeat if there exists a preference relation on a set of arguments. To summarise, a set of arguments A defeats another set of arguments B only if B has insufficient counter-arguments, or B has sufficient counter-arguments and for those counter-arguments, B is not more preferred to A. This notion of defeat allows us to distinguish between a successful or failed attack.

Definition 3.2.8. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and $A, B \subseteq AR$ be two sets of*

*arguments. A **def** B (denoted as **def**(A, B)) if and only if:*

$$\textbf{defensible}(A, B) \wedge (\textbf{overpowers}(A, B) \vee \textbf{preferable}(\Theta_a(A, B), \Theta_a(B, A))).$$

Note that this definition differs from that proposed by Bench-Capon [22], Kaci et al. [77,78] as well as Bourguet et al. [25].The first motivation is to retain the underlying attacks relation and only use the preferences when a tie needs to be broken.

Definition 3.2.9. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$, we define $defeats_{PAAF}$ to be the set of all ordered pairs $\langle A, B \rangle$, where $A, B \subseteq AR$, such that **def**(A, B) is true.*

Each ordered pairs in the defeats relation is directly correlated with the attacks relation. For each ordered pair in the defeats relation, the set of relevant attacking arguments is a subset of the domain of the attacks relation and the set of relevant arguments under-attack is a subset of the range of the attacks relation. This correlation is important as it disallow any arbitrary ordered pair to be included in the defeats relation.

Lemma 3.2.1. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, for all **def**$(A, B) \in defeats_{PAAF}$, $\Theta_a(A, B) \subseteq domain(attacks_{PAAF})$ and $\Theta_u(B, A) \subseteq range(attacks_{PAAF})$*

Proof:
By definition, $\Theta_a(A, B) = \{\alpha | \alpha \in A \wedge \exists \beta \in B \text{ s.t. } attacks_{PAAF}(\alpha, \beta)\}$ and $\Theta_u(A, B) = \{\alpha | \alpha \in A \wedge \exists \beta \in B \text{ s.t. } attacks_{PAAF}(\beta, \alpha)\}$. By definition, **def**$(A, B)$ is true if and only if **defensible**$(A, B) \wedge (\textbf{overpowers}(A, B) \vee \textbf{preferable}(\Theta_a(A, B), \Theta_a(B, A)))$ holds. Recall that **defensible**(A, B) holds if and only if $\Theta_a(B, A) \subseteq \Theta_u(B, A)$ is true and **overpowers**(A, B) holds if and only if $\Theta_a(A, B) \setminus \Theta_u(A, B) \neq \emptyset$ is true, therefore, **def**(A, B) is true if and only if $\Theta_a(B, A) \subseteq \Theta_u(B, A) \wedge (\Theta_a(A, B) \setminus \Theta_u(A, B) \neq \emptyset \vee (\Phi(\Theta_a(A, B)) \leq \Phi(\Theta_a(B, A))))$ holds. It follows that for all **def**$(A, B) \in defeats_{PAAF}$, $\Theta_a(A, B) \subseteq domain(attacks_{PAAF})$ and $\Theta_u(B, A) \subseteq range(attacks_{PAAF})$. □

The defeats relation is non-monotonic with respect to the increase in cardinality of the arguments. However, this definition generalises that of Dung [52], as trivially, if all arguments are equally preferred then the defeats relation is the superset to the attacks relation as defined in Dung [52].

Lemma 3.2.2. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, if $|V| = 1$ then $defeats_{PAAF} \supseteq \{attacks_{PAAF}\}$.*

Proof:

Let us assume that $defeats_{PAAF} \not\supseteq \{attacks_{PAAF}\}$. It follows that $\{attacks_{PAAF}\} \setminus defeats_{PAAF} \neq \emptyset$. This implies that there exists an assertion $attacks(\alpha, \beta) \in attacks_{PAAF}$ and that $\mathbf{def}(\{\alpha\}, \{\beta\}) \notin defeats_{PAAF}$. By definition, $\mathbf{def}(\{\alpha\}, \{\beta\}) \in defeats_{PAAF}$ if and only if $\mathbf{defensible}(\{\alpha\}, \{\beta\}) \wedge$ ($\mathbf{overpowers}(\{\alpha\}, \{\beta\}) \vee \mathbf{preferable}(\Theta_a(\{\alpha\}, \{\beta\}), \Theta_a(\{\alpha\}, \{\beta\})))$ is true. Thus entails that either the defensible condition or overpower and preferable conditions are violated. Let us consider two cases:

Case 1: Violation of the defensible condition (i.e. $\neg\mathbf{defensible}(\{\alpha\}, \{\beta\})$). By definition $\mathbf{defensible}$(A,B) if and only if $\Theta_a(B, A) \subseteq \Theta_u(B, A)$ is true. Recall that $\Theta_a(A, B) = \{\alpha | \alpha \in A \wedge \exists \beta \in B \text{ s.t. } attacks_{PAAF}(\alpha, \beta)\}$ and $\Theta_u(A, B) = \{\alpha | \alpha \in A \wedge \exists \beta \in B \text{ s.t. } attacks_{PAAF}(\beta, \alpha)\}$, this entails that for each $attacks(\alpha, \beta) \in attacks_{PAAF}$, $\mathbf{defensible}(\{\alpha\}, \{\beta\})$ is true, thus, violating the assumption.

Case 2: Violation of the overpower and preferable condition (i.e. $\neg\mathbf{overpowers}(A, B) \wedge \neg\mathbf{preferable}(\Theta_a(A, B), \Theta_a(B, A))))$. Given $|V| = 1$ and by definition, Φ is a total valuation function, this entails that all arguments are equally preferred. Therefore, $\mathbf{preferable}(\Theta_a(A, B), \Theta_a(B, A))$ is always true, thus, violating the assumption. $\qquad\square$

Given the lemma 3.2.2, it naturally follows that if the set of arguments and $attacks_{PAAF}$ relationship in an abstract preference-based accrual argumentation framework is identical to that defined in an abstract argumentation framework as defined in Dung [52], then the $defeats_{PAAF}$ relation is the superset to the $attacks_{AF}$ relation.

Theorem 3.2.3. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, an abstract argumentation framework $AF = \langle AR, attacks_{AF} \rangle$ as in the definition of Dung [52], and $attacks_{PAAF} = attacks_{AF}$, if $|V| = 1$ then $defeats_{PAAF} \supseteq \{attacks_{AF}\}$.*

Proof:

Lemma 3.2.2 shows that $defeats_{PAAF} \supseteq \{attacks_{PAAF}\}$. Since the $attacks_{PAAF} = attacks_{AF}$ then it follows that $defeats_{PAAF} \supseteq \{attacks_{AF}\}$. $\qquad\square$

In the following definition, we introduce the notion of a maximal defeats relation. This notion of a maximal defeats relation is unique to $PAAF$. Since the defeats relation is non-monotonic with respect to the accrued sets of argument, the aim of a maximal defeat relation identifies the subset of defeats relations that have reached a fix-point. In line with the

principle highlighted by Prakken [122], where arguments involved in an accrual cannot be considered individually once an accrual has been performed, we utilise the maximal defeats relation to ignore defeats relations where accrual of the arguments involved in the defeats relation has occurred.

Definition 3.2.10. *Given a* $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ *and* $A, B \subseteq AR$ *be two sets of conflict-free arguments. If* $\textbf{def}(A, B)$*,* A ***maximally defeats*** B *(denoted as* $\textbf{def}_M(A, B)$*) if and only if:*

$$\neg \exists A'(A' \text{ is conflict-free} \wedge (A \subset A' \subseteq AR \wedge \textbf{def}(B, A'))).$$

Given the definition of defeats and maximal defeats, let us consider the following example.

Example 3.2.3. *Continuing from example 3.2.2, we will assume the existence of a set of abstract values* $\{v_1, v_2, v_3, v_4\}$ *with a total ordering on the abstract values represented as* $\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$*. From the initial example, we know that Tom has the goal to be fit and believed that the combination of heat and rain may be less unpleasant than heat or rain alone. In such a situation, the accrual of arguments weaken the conclusive strength. Let us assume that* Φ *provides the following valuations: both* $\{\beta\}, \{\gamma\}$ *are assigned* v_1*,* $\{\alpha\}$ *is assigned* v_2*,* $\{\beta, \gamma\}$ *are assigned* v_3 *and all other combinations are assigned* v_4*. Table 3.3 presents a summary of this discussion. By using the set of values, and the provided valuation function* Φ*, we construct the defeat and maxi-defeat relations. Utilising digraphs, figure 3.2 presents a graphical illustration of the defeat interaction between the arguments. Arguments are represented as labelled vertices and the* $attacks_{PAAF}$ *relation as directed edges. Hence* $attacks_{PAAF}(\alpha, \beta)$ *is represented with a directed edge from the vertex* α *to* β*.*

Now that we have established a basic definition for the abstract framework, we will require a notion of acceptability. Our notion of acceptability is defined with respect to a set of arguments. A set of arguments A is acceptable to another set of arguments S (in other words accepted into the set) if the set of arguments S defends A from any defeat.

Definition 3.2.11. *Given a* $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ *and* $A \subseteq AR$*.* A *is* ***acceptable*** *with respect to a set of arguments* $S \subseteq AR$ *(denoted as* $\textbf{acceptable}(A, S)$*) if and only if:*

$$\forall B((B \subseteq AR) \wedge \textbf{def}_M(B, A) \rightarrow \textbf{def}_M(S, B)).$$

Note that this definition differs from that proposed by Dung [52] as well as Bench-Capon [22]. We have defined the acceptability based on the acceptance of a set of arguments to

(a) Exchange 1

(b) Exchange 2

(c) Exchange 3
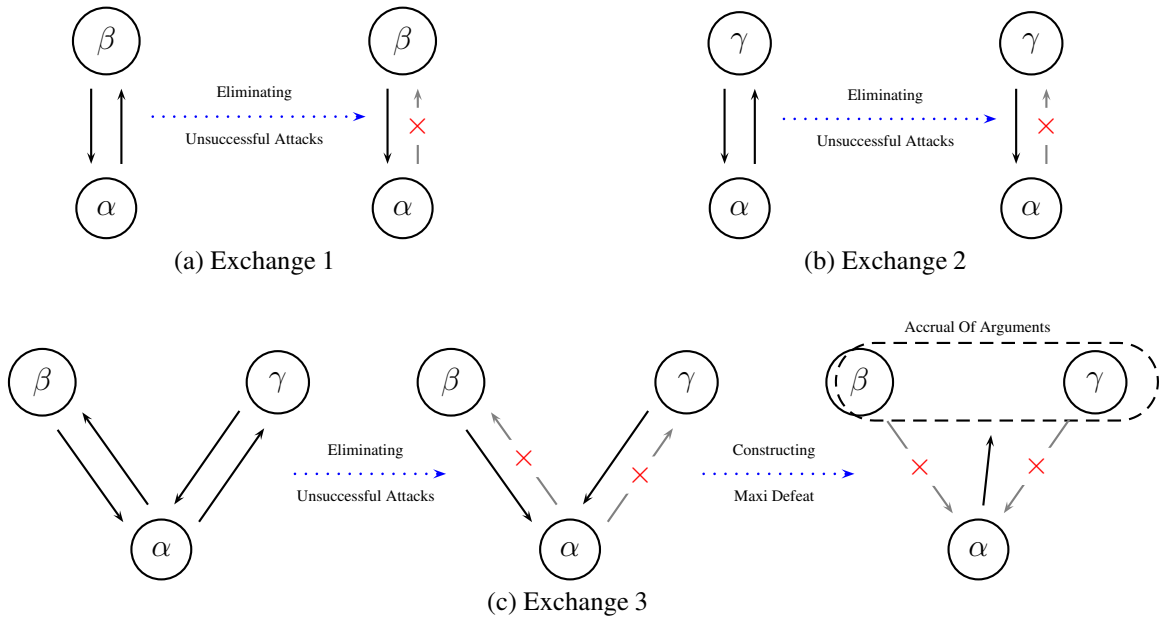
In figure 3.2a, figure 3.2b and figure 3.2c, arguments are represented by the vertices α, β and γ. The directed edges represent the relations $attacks_{PAAF}(\alpha, \beta)$ and $attacks_{PAAF}(\beta, \alpha)$, $attacks_{PAAF}(\alpha, \gamma)$ and $attacks_{PAAF}(\gamma, \alpha)$, and $attacks_{PAAF}(\alpha, \beta)$, $attacks_{PAAF}(\beta, \alpha)$, $attacks_{PAAF}(\alpha, \gamma)$ and $attacks_{PAAF}(\gamma, \alpha)$ respectively. The $defeat_{PAAF}$ relations represent successful attacks, hence, unsuccessful attacks are eliminated. The eliminated directed edges are marked with a cross. An accrual of arguments is represented by a dotted circle encapsulating the accrued arguments. Hence, figure 3.2a is interpreted as: in the instance where α is attacking β and β is attacking α, β defeats α. Figure 3.2b is interpreted as: in the instance where α is attacking γ and γ is attacking α, γ defeats α. Figure 3.2c is interpreted as: in the instance where α is attacking β and γ while both β and γ are attacking α separately, α is defeated by β and γ. However, if β and γ are accrued, the resulting accrued set of arguments ($\{\beta, \gamma\}$) is defeated by α.

Figure 3.2: Graphical Illustration of Defeat

another set rather than a single argument. However, this definition generalised that of Dung [52], as trivially, a singleton set will capture the same notion as defined in Dung [52].

Given a notion of conflict-freedom and acceptability, we will now provide a notion of admissibility. Admissibility can simply be defined as an attribute of a set of arguments that is conflict-free and that defends itself from all defeats.

Definition 3.2.12. *Given a* $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ *and a conflict-free set of arguments* $S \subseteq AR$ *is **admissible** if and only if:*

$$acceptable(S, S)$$

	Exchange 1	Exchange 2	Exchange 3
AR	$\{\alpha, \beta\}$	$\{\alpha, \gamma\}$	$\{\alpha, \beta, \gamma\}$
$attacks_{PAAF}$	$attacks_{PAAF}(\alpha, \beta)$ $attacks_{PAAF}(\beta, \alpha)$	$attacks_{PAAF}(\gamma, \alpha)$ $attacks_{PAAF}(\alpha, \gamma)$	$attacks_{PAAF}(\alpha, \beta), attacks_{PAAF}(\beta, \alpha)$ $attacks_{PAAF}(\gamma, \alpha), attacks_{PAAF}(\alpha, \gamma)$
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}$	$\emptyset, \{\alpha\}, \{\gamma\}$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
V		$\{v_1, v_2, v_3, v_4\}$	
$\leq$		$\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$	
Φ	$\Phi(\{\beta\}) = v_1,$ $\Phi(\{\alpha\}) = v_2,$ $\Phi(\{\alpha, \beta\}) = v_4$ $\Phi(\emptyset) = v_4$	$\Phi(\{\gamma\}) = v_1,$ $\Phi(\{\alpha\}) = v_2,$ $\Phi(\{\alpha, \gamma\}) = v_4$ $\Phi(\emptyset) = v_4$	$\Phi(\{\beta\}) = v_1, \Phi(\{\beta, \gamma\}) = v_3,$ $\Phi(\{\gamma\}) = v_1, \Phi(\{\alpha, \beta\}) = v_4,$ $\Phi(\{\alpha\}) = v_2, \Phi(\{\alpha, \gamma\}) = v_4,$ $\Phi(\emptyset) = v_4, \Phi(\{\alpha, \beta, \gamma\}) = v_4$
$defeats_{PAAF}$	$\mathbf{def}(\{\beta\}, \{\alpha\})$	$\mathbf{def}(\{\gamma\}, \{\alpha\})$	$\mathbf{def}(\{\beta\}, \{\alpha\}),$ $\mathbf{def}(\{\gamma\}, \{\alpha\}),$ $\mathbf{def}(\{\alpha\}, \{\beta, \gamma\})$
maxi-defeats	$\mathbf{def_M}(\{\beta\}, \{\alpha\})$	$\mathbf{def_M}(\{\gamma\}, \{\alpha\})$	$\mathbf{def_M}(\{\alpha\}, \{\beta, \gamma\})$

This table provides a summary of the discussion in example 3.2.3 and illustrates how preferences are presented in $PAAF$. It also demonstrates the assignment of arguments to preference values by the valuation function Φ as well as the computation of the defeat and maximally defeat relation.

Table 3.3: Summary of Example 3.2.3

Given the definition of acceptability and admissibility, let us consider the following example.

Example 3.2.4. *Continuing from Example 3.2.3, the notion of admissibility consists of two components. Firstly, the set of arguments must be conflict-free. Secondly, all subsets of the set of arguments must be acceptable to the set. With this in mind, we can further filter the conflict-free sets of arguments to those that are admissible. The following admissible set exists: The empty-set remains admissible in all three exchanges since it is always conflict-free and acceptable to itself. In exchange 1, $\{\beta\}$ is conflict-free and $\{\beta\}$ maximally defeats $\{\alpha\}$. Therefore, the additional admissible set is $\{\beta\}$. Similarly, in exchange 2, $\{\gamma\}$ is conflict-free and $\{\gamma\}$ maximally defeats $\{\alpha\}$. Hence, the additional admissible set is $\{\gamma\}$. In exchange 3, although $\{\beta, \gamma\}$ is conflict-free, $\{\beta, \gamma\}$ is maximally defeated by $\{\alpha\}$. Therefore, $\{\alpha\}$ exists as the other admissible set of arguments in exchange 3. Table 3.4 presents a summary of this discussion.*

In any argumentation system, it is conceivable that there are many potential admissible sets of arguments. For example, given two arguments (α, β) and the two defeats relation assertions ($\mathbf{def}(\{\alpha\}, \{\beta\})$, $\mathbf{def}(\{\beta\}, \{\alpha\})$), both singleton sets $\{\alpha\}$ and $\{\beta\}$ are admissible sets of arguments. Given a set of admissible sets of arguments, one might wish to order these admissible sets. Ordering can be performed with respect to maximal set inclusion such as the definition of Dung [52] or with respect to preferences such as the definition of Bench-Capon [22] and Bourguet et al. [25]. Taking the resulting ordering of these two approaches as

	Exchange 1	Exchange 2	Exchange 3
AR	$\{\alpha, \beta\}$	$\{\alpha, \gamma\}$	$\{\alpha, \beta, \gamma\}$
$attacks_{PAAF}$	$attacks_{PAAF}(\alpha, \beta)$ $attacks_{PAAF}(\beta, \alpha)$	$attacks_{PAAF}(\gamma, \alpha)$ $attacks_{PAAF}(\alpha, \gamma)$	$attacks_{PAAF}(\alpha, \beta), attacks_{PAAF}(\beta, \alpha)$ $attacks_{PAAF}(\gamma, \alpha), attacks_{PAAF}(\alpha, \gamma)$
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}$	$\emptyset, \{\alpha\}, \{\gamma\}$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
V		$\{v_1, v_2, v_3, v_4\}$	
$\leq$		$\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$	
Φ	$\Phi(\{\beta\}) = v_1,$ $\Phi(\{\alpha\}) = v_2,$ $\Phi(\{\alpha, \beta\}) = v_4$ $\Phi(\emptyset) = v_4$	$\Phi(\{\gamma\}) = v_1,$ $\Phi(\{\alpha\}) = v_2,$ $\Phi(\{\alpha, \gamma\}) = v_4$ $\Phi(\emptyset) = v_4$	$\Phi(\{\beta\}) = v_1, \Phi(\{\beta, \gamma\}) = v_3,$ $\Phi(\{\gamma\}) = v_1, \Phi(\{\alpha, \beta\}) = v_4,$ $\Phi(\{\alpha\}) = v_2, \Phi(\{\alpha, \gamma\}) = v_4,$ $\Phi(\emptyset) = v_4, \Phi(\{\alpha, \beta, \gamma\}) = v_4$
$defeats_{PAAF}$	**def**$(\{\beta\}, \{\alpha\})$	**def**$(\{\gamma\}, \{\alpha\})$	**def**$(\{\beta\}, \{\alpha\}),$ **def**$(\{\gamma\}, \{\alpha\}),$ **def**$(\{\alpha\}, \{\beta, \gamma\})$
maxi-defeats	**def$_\mathbf{M}$**$(\{\beta\}, \{\alpha\})$	**def$_\mathbf{M}$**$(\{\gamma\}, \{\alpha\})$	**def$_\mathbf{M}$**$(\{\alpha\}, \{\beta, \gamma\})$
admissible	$\emptyset, \{\beta\}$	$\emptyset, \{\gamma\}$	$\emptyset, \{\alpha\}$

This table provides a summary of the discussion in example 3.2.4 and illustrates the computation of the admissible sets of arguments.

Table 3.4: Summary of Example 3.2.4

two extremes of a spectrum, we propose an approach that incorporate both these approaches hence providing a range of possible combinations within this spectrum. In the proposed approach, we firstly apply the preferences to constraint the set of admissible sets of arguments. The resulting set of admissible sets of arguments is then constrained further by removing admissible sets that are not maximal with respect to set inclusion. By varying the preference placed on the admissible sets, we are able to change where the resulting outcome lays within the spectrum of possible ordering combination. The next two definitions allow us to order the set of admissible sets. The ordering on the set of admissible set is induced by mapping each admissible set to a value in the set of abstract values. Each most preferred admissible set constitutes a set of accrued arguments. Hence, we described these as "accrued" extensions.

Definition 3.2.13. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and an admissible set of arguments $S \subseteq AR$, S is an **accrued extension** if there is no admissible set $S' \subseteq AR$ such that $pref(\Phi(S'), \Phi(S))$.*

Following the approach of Dung [52], preferred extensions are admissible sets that are maximal with respect to set inclusion. However, our notion of preferred extension needs to capture the notion of accrual. Hence, we will only consider admissible sets from the set of accrued extensions.

Definition 3.2.14. *Given a $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ and an accrued extension $S \subseteq AR$, S is a **preferred extension** if there is no accrued extension $S' \subseteq AR$ such that $S \subset S'$.*

Note that the notion of preferred extension as the definition of Dung [52] is a special case of the definition given above.

Theorem 3.2.4. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$ and an abstract argumentation framework $AF = \langle AR, attacks_{AF} \rangle$ as in the definition of Dung [52], if $|V|$ = 1 then an admissible set in $PAAF$ is also an admissible set in AF.*

Proof:

Recall that a set of arguments S is admissible in AF if and only if the set S is conflict-free and all arguments in S is acceptable to S. Assume that the set S is admissible in $PAAF$ but not admissible in AF, let us consider two cases:

Case 1: S is conflict-free in $PAAF$ and not conflict-free in AF. Given that the definition of conflict-freedom in $PAAF$ is the same as that defined in AF, then for S to be conflict-free in $PAAF$ and not conflict-free in AF entails that there exists an argument in S that attacks another argument in S with respect to AF. It then follows that there exists an assertion in the set of attacks relation in AF that is not in the set of attacks relation in $PAAF$. However, since the attacks relation in $PAAF$ is the same as the attacks relation in AF, it cannot be the case that an argument in S that attacks another argument in S in AF and not in $PAAF$, hence, violating the assumption.

Case 2: S is acceptable in $PAAF$ and not acceptable in AF. Recall that a set of arguments S is acceptable in AF if any arguments attacking S is attacked by some arguments in S. If S is not acceptable in AF, it entails that there exists an argument $\alpha \in AR$ that attacks some argument in S and is not attacked by S[1]. Since the set of attacks relation in $PAAF$ is equal to the set of attacks relation in AF, it follows that there also exists an argument $\alpha \in AR$ in $PAAF$ that attacks some argument in S and is not attacked by S. However, for S to be acceptable, it must be the case that S maximally defeats α. Given that $|V|$ = 1, lemma 3.2.2 shows that $defeats_{PAAF} \supseteq \{attacks_{PAAF}\}$, lemma 3.2.1 shows that each assertion in $defeats_{PAAF}$ is constructed from assertions from $attacks_{PAAF}$ and $\mathbf{def_M} \subset defeats_{PAAF}$, it cannot be the case that S maximally defeats α. Hence it cannot be the case that S is acceptable in $PAAF$, thus, violating the assumption. □

---

[1] as utilised by Dung [52] to mean attacked by some argument in S

Theorem 3.2.5. *Given an abstract preference-based accrual argumentation framework* $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ *where* $Bel = \langle V, \leq, \Phi \rangle$ *and an abstract argumentation framework* $AF = \langle AR, attacks_{PAAF} \rangle$ *as in the definition of Dung [52], if* $|V| = 1$ *then a preferred extension in* $PAAF$ *is also a preferred extension in* AF.

Proof:

Given that $|V| = 1$, theorem 3.2.4 shows that admissible sets in $PAAF$ are also admissible sets in AF. By the definition of accrued extension, if $|V| = 1$, then all admissible sets of arguments are equally preferred. Therefore, all admissible sets of arguments are accrued extension. By definition, a preferred extension is a maximal (with respect to set inclusion) accrued extension. Recall the definition of Dung [52], a preferred extension is a maximal (with respect to set inclusion) admissible set of arguments. Hence a preferred extension in $PAAF$ is a preferred extension in AF $\qquad\qquad\square$

Given the definition of accrued and preferred extension, let us consider the following example.

Example 3.2.5. *Continuing from Example 3.2.4, given the conflict-free sets* $\{\emptyset, \{\alpha\}, \{\beta\}\}$, $\{\emptyset, \{\alpha\}, \{\gamma\}\}$, $\{\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}\}$ *for exchange 1,2 and 3 respectively, and the associated* Φ *with an ordering of* $\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$ *on the values, we can determine the accrued and preferred extensions. In exchange 1, the accrued and preferred extension is* $\{\beta\}$. *In exchange 2, the accrued and preferred extension is* $\{\gamma\}$. *In exchange 3, the accrued and preferred extension is* $\{\alpha\}$. *Table 3.5 presents a summary of this discussion.*

As highlighted by studies as such [75, 127, 135], the use of argumentation in negotiation can be viewed as a distributed search process over a space consisting of potential solutions. When searching for potential solutions in domains such as legal judgements or decision support, an argumentation system might be called upon to provide a definitive answer. In such a situation, the generation of a unique extension is desirable. However, when argumentation systems are deployed for what-if analysis, the generation and exploration of competing alternatives are required. In such a situation, the generation of multiple extensions is desirable. Ideally, an argumentation framework should accommodate both abilities of generating a unique extension or multiple extensions depending on the given situation. Let us first focus on conditions required to generate a unique extension in $PAAF$. There are two distinct conditions where a unique extension can be generated:

- There is no cycle in the attacks relation.

	Exchange 1	Exchange 2	Exchange 3
AR	$\{\alpha, \beta\}$	$\{\alpha, \gamma\}$	$\{\alpha, \beta, \gamma\}$
$attacks_{PAAF}$	$attacks_{PAAF}(\alpha, \beta)$ $attacks_{PAAF}(\beta, \alpha)$	$attacks_{PAAF}(\gamma, \alpha)$ $attacks_{PAAF}(\alpha, \gamma)$	$attacks_{PAAF}(\alpha, \beta), attacks_{PAAF}(\beta, \alpha)$ $attacks_{PAAF}(\gamma, \alpha), attacks_{PAAF}(\alpha, \gamma)$
conflict-free	$\emptyset, \{\alpha\}, \{\beta\}$	$\emptyset, \{\alpha\}, \{\gamma\}$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$
V		$\{v_1, v_2, v_3, v_4\}$	
$\leq$		$\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$	
Φ	$\Phi(\{\beta\}) = v_1,$ $\Phi(\{\alpha\}) = v_2,$ $\Phi(\{\alpha, \beta\}) = v_4,$ $\Phi(\emptyset) = v_4$	$\Phi(\{\gamma\}) = v_1,$ $\Phi(\{\alpha\}) = v_2,$ $\Phi(\{\alpha, \gamma\}) = v_4$ $\Phi(\emptyset) = v_4$	$\Phi(\{\beta\}) = v_1, \Phi(\{\beta, \gamma\}) = v_3,$ $\Phi(\{\gamma\}) = v_1, \Phi(\{\alpha, \beta\}) = v_4,$ $\Phi(\{\alpha\}) = v_2, \Phi(\{\alpha, \gamma\}) = v_4,$ $\Phi(\emptyset) = v_4, \Phi(\{\alpha, \beta, \gamma\}) = v_4$
$defeats_{PAAF}$	$\mathbf{def}(\{\beta\}, \{\alpha\})$	$\mathbf{def}(\{\gamma\}, \{\alpha\})$	$\mathbf{def}(\{\beta\}, \{\alpha\}),$ $\mathbf{def}(\{\gamma\}, \{\alpha\}),$ $\mathbf{def}(\{\alpha\}, \{\beta, \gamma\})$
maxi-defeats	$\mathbf{def_M}(\{\beta\}, \{\alpha\})$	$\mathbf{def_M}(\{\gamma\}, \{\alpha\})$	$\mathbf{def_M}(\{\alpha\}, \{\beta, \gamma\})$
admissible	$\emptyset, \{\beta\}$	$\emptyset, \{\gamma\}$	$\emptyset, \{\alpha\}$
accrued	$\{\beta\}$	$\{\gamma\}$	$\{\alpha\}$
preferred	$\{\beta\}$	$\{\gamma\}$	$\{\alpha\}$

This table provides a summary of the discussion in example 3.2.5 and illustrates the computation of the accrued and preferred sets of arguments.

Table 3.5: Summary of Example 3.2.5

- Each accrual set of arguments is mapped to a unique preference value.

Firstly, if there is no cycle in the attacks relation, the preferences are never utilised in the construction of the defeats relation. As shown in lemma 3.2.2, the defeats relation is a superset of the attacks relation. Therefore, there exists exactly one preferred extension. No cycles in the attack relationship essentially places a total ordering on the arguments forming a chain. The preferred extension hence consists of the argument that is not attacked by any argument and all alternate arguments in the chain as shown by Bench-Capon [22]. If there exist cycles in the attacks relation then the unique extension generation condition depends on the preferences order. In such a situation, the condition that each accrual set of arguments is mapped to a unique preference value will produce a unique extension. This additional condition can be achieved by utilising an injective valuation function. Using such a valuation function allows arguments to be assigned a unique value hence in $PAAF$, a total ordering will be induced on the set of arguments.

Lemma 3.2.6. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, if Φ is injective and a relation $\preceq$ is defined on 2^{AR} s.t. $\forall AR_1, AR_2 \in 2^{AR}, AR_1 \preceq AR_2$ if $\Phi(AR_1) \leq \Phi(AR_2)$, then $\preceq$ is a total order.*

Proof:

$\preceq$ is a total order if and only if it satisfies the following properties:

1. Anti-symmetry: Let us assume that there exists $a, b, \in 2^{AR}$ s.t. $a \preceq b$, $b \preceq a$ and $a \neq b$. $a \preceq b$ entails that $\Phi(a) \leq \Phi(b)$. $b \preceq a$ entails that $\Phi(b) \leq \Phi(a)$. Because Φ is injective, $a \neq b$ entails $\Phi(a) \neq \Phi(b)$. Therefore, we have $\Phi(a) \leq \Phi(b)$, $\Phi(b) \leq \Phi(a)$ and $\Phi(a) \neq \Phi(b)$, which entails that $\leq$ violates the property of anti-symmetry (recall $\leq$ is a total order). Hence the assumption is invalid.

2. Transitivity: Let us assume that there exists $a, b, c \in 2^{AR}$ s.t $a \preceq b$, $b \preceq c$ and $a \npreceq c$. $a \preceq b$ entails that $\Phi(a) \leq \Phi(b)$. $b \preceq c$ entails that $\Phi(b) \leq \Phi(c)$. Because $\leq$ is a total order and must satisfy the property of transitivity, it must be the case that $\Phi(a) \leq \Phi(c)$. By our construction of $\preceq$, this entails that $a \preceq c$, which violates our assumption.

3. Totality: Let us assume that there exists $a, b, \in 2^{AR}$ s.t. $a \npreceq b$ and $b \npreceq a$. Let us consider two cases:

Case 1: $a = b$. Then by definition, $\Phi(a) = \Phi(b)$. Recall $\leq$ is a total order. By the totality property of $\leq$, we have either $\Phi(a) \leq \Phi(b)$ or $\Phi(b) \leq \Phi(a)$. Then, by our construction of $\preceq$, it must be the case that $a \preceq b$ or $b \preceq a$. Hence, the assumption is violated.

Case 2: $a \neq b$. Because Φ is injective, $a \neq b$ entails $\Phi(a) \neq \Phi(b)$. Recall $\leq$ is a total order. By definition, a total order is anti-symmetric. Hence, $\Phi(a) \neq \Phi(b)$ entails that we have not $\Phi(a) \leq \Phi(b)$ and $\Phi(b) \leq \Phi(a)$. By the totality property of $\leq$, we have either $\Phi(a) \leq \Phi(b)$ or $\Phi(b) \leq \Phi(a)$. Therefore, we have either $\Phi(a) \leq \Phi(b)$ and $\Phi(b) \nleq \Phi(a)$ or $\Phi(a) \nleq \Phi(b)$ and $\Phi(b) \leq \Phi(a)$. By our construction of $\preceq$, this entails that either $a \preceq b$ and $b \npreceq a$ or $a \npreceq b$ and $b \preceq a$, thus violating the assumption. $\square$

It is interesting to note that the inclusion relation $\subseteq$ defines a partial order on a power-set. However, as shown in examples 3.1.1 and 3.2.5, the accrued arguments may sometime be less preferred to individual arguments resulting in the superset of arguments being less preferred to the subset of arguments. Therefore, $\preceq$ is not the inclusion relation. To generate a total ordering on the arguments, an additional condition on the values must exist. The cardinality of the set of values must be greater or equal to the cardinality of the set of arguments.

Proposition 3.2.7. *Given an abstract preference-based accrual argumentation framework* $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ *where* $Bel = \langle V, \leq, \Phi \rangle$, *if* Φ *is injective then* $|2^{AR}| \leq |V|$.

Proof:

Let us assume that $|V| < |2^{AR}|$. Given that Φ is injective and by definition Φ is a total valuation function, Φ is a total injective valuation function. This implies that $\forall a \in 2^{AR}, \exists v_1 \in V$

s.t $(a, v_1) \in \Phi$ and $\forall a, b \in 2^{AR}, v_1, v_2 \in V$ if $(a, v_1), (b, v_2) \in \Phi$ then $a \neq b$ and $v_1 \neq v_2$. Therefore there exists as many elements in 2^{AR} as in V, thus, violating the assumption. $\square$

The following theorems show that, given the described conditions, the formal system generates a unique extension. In other words, there exists exactly one accrued and preferred extension.

Theorem 3.2.8. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, if Φ is injective and a relation $\preceq$ is defined on 2^{AR} s.t. $\forall AR_1, AR_2 \in 2^{AR}, AR_1 \preceq AR_2$ if $\Phi(AR_1) \leq \Phi(AR_2)$, then $\preceq$ is a total order on the set of admissible sets of arguments (with respect to $PAAF$).*

Proof:
Let us denote AS to represent the set of admissible sets of arguments in $PAAF$. Lemma 3.2.6 shows that $\preceq$ is a total order on 2^{AR}. We note that $AS \subseteq 2^{AR}$. Therefore $\preceq$ is also a total order on AS. $\square$

Theorem 3.2.9. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, there exists at least one accrued extension.*

Proof:
Let us assume that there does not exist any accrued extension. This entails that there does not exist any admissible sets of arguments. However, the empty-set is always admissible, thus violating the assumption. $\square$

Theorem 3.2.10. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, if Φ is injective then there exists exactly one accrued extension and the accrued extension is also preferred.*

Proof:
Assume that $S_1, S_2 \subseteq AR$ are *accrued extension* and that $S_1 \neq S_2$. Because Φ is injective, $S_1 \neq S_2$ implies $\Phi(S_1) \neq \Phi(S_2)$. By virtue of the fact that $\leq$ is a total order, it must be the case that $\Phi(S_1) \leq \Phi(S_2)$ or $\Phi(S_2) \leq \Phi(S_1)$. By the definition of an accrued extension, if $\Phi(S_1) \leq \Phi(S_2)$ then S_2 is not an accrued extension and if $\Phi(S_2) \leq \Phi(S_1)$ then S_1 is not an accrued extension, thus violating the assumptions. If Φ is injective, then there is only one accrued extension. By the definition of a preferred extension, this accrued extension is also a preferred extension. $\square$

The ability of an argumentation framework to generate a unique solution is an important property for use in situations such as automated reasoning, negotiation, argumentation and

in group decision support. In essence, the ability to identify the unique solution entails that the argumentation system is able to provide a definitive answer where one exists. This allows for the use of argumentation technology as the decision making machinery in an agent-based system, either internalised within the agents or as a protocol for achieving collaborative goal in multi-agent systems.

Let us now consider the conditions required for generating multiple extensions. Multiple extensions entail that there is no definitive answer but rather several competing and equally plausible alternatives. To achieve this, the following two conditions must exist:

- There are cycles in the attacks relation.

- Each preferences value is associated with more than one argument.

Firstly, there must exist at least two arguments that are attacking each other, forming a cycle in the attacks relation. Furthermore, ties are not broken by the use of preferences. By having more than one argument associated with each preference value, a total pre-order is formed on the set of arguments. Therefore, unresolved cycles forms the basis for multiple extension. This additional condition can be achieved by utilising a surjective valuation function. Using such a valuation function allows for multiple arguments to be assigned the same value hence in $PAAF$, a total pre-order will be induced on the set of arguments.

Theorem 3.2.11. *Given an abstract preference-based accrual argumentation framework $PAAF = \langle AR, attacks_{PAAF}, Bel \rangle$ where $Bel = \langle V, \leq, \Phi \rangle$, if Φ is surjective and a relation $\preceq$ is defined on 2^{AR} s.t. $\forall AR_1, AR_2 \in 2^{AR}, AR_1 \preceq AR_2$ if $\Phi(AR_1) \leq \Phi(AR_2)$, then $\preceq$ is a total pre-order.*

Proof:
$\preceq$ is a total pre-order if and only if it satisfies the following properties:

1. Transitivity: Let us assume that there exist $a, b, c \in 2^{AR}$ s.t $a \preceq b$, $b \preceq c$ and $a \not\preceq c$. $a \preceq b$ entails that $\Phi(a) \leq \Phi(b)$. $b \preceq c$ entails that $\Phi(b) \leq \Phi(c)$. Because $\leq$ is a total order and must satisfy the property of transitivity, it must be the case that $\Phi(a) \leq \Phi(c)$. By our construction of $\preceq$, this entails that $a \preceq c$, which violates our assumption.

2. Totality: Let us assume that there exist $a, b, \in 2^{AR}$ s.t. $a \not\preceq b$ and $b \not\preceq a$. Let us consider two cases:

Case 1: $a = b$. Recall Φ is a total function, then, by definition $\Phi(a)$ and $\Phi(b)$ are valid valuations. Recall $\leq$ is a total order, by the totality property of $\leq$, we have either

$\Phi(a) \leq \Phi(b)$ or $\Phi(b) \leq \Phi(a)$. Then, by our construction of $\preceq$, it must be the case that $a \preceq b$ or $b \preceq a$. Hence, the assumption is violated.

Case 2: $a \neq b$. Recall Φ is a total function, then, by definition $\Phi(a)$ and $\Phi(b)$ are valid valuation. Recall $\leq$ is a total order, by the totality property of $\leq$, we have either $\Phi(a) \leq \Phi(b)$ or $\Phi(b) \leq \Phi(a)$. Then, by our construction $\preceq$, this entails that either $a \preceq b$ or $b \preceq a$, thus violating the assumption. $\square$

The ability of an argumentation framework to generate multiple competing but plausible solutions is an important property for use in situations such as exploratory analysis, decision solution discovery. This allows for the use of argumentation technology in scenario discovery and group decision support tools.

In following section, we will firstly present a brief comparison between AF, VAF and $PAAF$. Secondly, perform a discussion on the utilisation of preferences as informed representation of argument sources, credibility and trust.

3.3 Discussion

In the previous section, we introduced the abstract preference-based accrual argumentation framework ($PAAF$) and illustrated the ability of the framework to perform accrual of arguments and highlighted some of the unique features of the framework. Our aims in this section are two-fold: Firstly, we aim to present a discussion and comparison illustrating the differences between the abstract argumentation framework (AF), the value-based argumentation framework (VAF) and $PAAF$. Secondly, we aim to address the use of preferences as informed representation of argument sources, credibility and trust such as those situations illustrated in the motivating example 1.2.3 and 1.2.4.

3.3.1 Comparison Between Frameworks

To make a comparison between AF, VAF and $PAAF$, we will consider two examples (example 3.1.1 and 1.2.2). Let us first consider the motivating example 3.1.1. As one might recall, in example 3.1.1, Tom was debating if he should go jogging. Three arguments: "Tom has not been jogging for several days, so he should go jogging.", "It is raining, so Tom should not go jogging." and "It is hot, so Tom should not go jogging." were presented. In

the example, Tom decided that he should go jogging because the combination of rain and the heat actually weakens the arguments for not jogging. This example shows that in certain situations, the accrual of arguments can result in a weaker position.

Let us consider the example formulated in AF, VAF and $PAAF$.

Example 3.3.1. *Let us assume that $\{\alpha, \beta, \gamma\}$ are arguments representing: "Tom has not been jogging for several days, so he should go jogging.", "It is raining, so Tom should not go jogging." and "It is hot, so Tom should not go jogging." respectively. The attacks relationship capturing the interplay between the arguments are represented as $\{attacks(\alpha, \beta),$ $attacks(\beta, \alpha),$ $attacks(\alpha, \gamma),$ $attacks(\gamma, \alpha)\}$. Given the set of arguments and attack relation, the conflict-free sets of arguments for AF, VAF and $PAAF$ are $\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$. In other words, the set of conflict-free sets of arguments is the same for all three frameworks.*

Let us now assume the existence of a set of abstract values $\{v_1, v_2, v_3, v_4\}$ with a total ordering on the abstract values represented as $\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$ and a function Φ that maps the arguments to their respective preference values. From the motivating example 3.1.1, we know that Tom has the goal to be fit and believes that the combination of heat and rain may be less unpleasant than heat or rain alone. In such a situation, the accrual of arguments weakens the strength of set of accrued arguments. To reflect Tom's preferences, we will assign v_1 to both $\{\beta\}$ and $\{\gamma\}$, v_2 to $\{\alpha\}$, v_3 to $\{\beta, \gamma\}$ and v_4 to all other combinations. Table 3.6 presents a summary of the outcomes for each framework. Note that the preferred extension for AF and VAF is $\{\beta, \gamma\}$ while $\{\alpha\}$ is the preferred extension for $PAAF$. This illustrates that both AF and VAF are not able to capture the intention highlighted in the example.

Let us now consider the motivating example 1.2.2, where a debate between two people arguing whether a particular person (Bill) should be incarcerated was presented. In this situation, one person argues that he should because of a crime he has committed, while the other argues that he should not because of his age. The three arguments "Bill is a juvenile; therefore he should not go to jail.", "Bill has assaulted someone, therefore he should be jailed." and "Bill has robbed someone, therefore he should be jailed." were presented and the conclusion was that Bill should go to jail. This example illustrates that in certain situations, the accrual of arguments can result in a stronger position.

Let us consider the example formulated in AF, VAF and $PAAF$.

Example 3.3.2. *Assume that $\{\alpha, \beta, \gamma\}$ are arguments representing: "Bill is a juvenile; therefore he should not go to jail.", "Bill has assaulted someone, therefore he should be jailed."*

	AF	VAF	$PAAF$
AR		$\{\alpha, \beta\,\gamma\}$	
attacks		$\{attack(\alpha,\gamma), attack(\alpha,\beta), atttack(\beta,\alpha), attack(\gamma,\alpha)\}$	
conflict-free		$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$	
V	$-$	$\{v_1, v_2, v_3, v_4\}$	
$\leq$	$-$	$\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$	
Φ	$-$	$\Phi_{VAF}(\gamma) = v_1,$ $\Phi_{VAF}(\beta) = v_1,$ $\Phi_{VAF}(\alpha) = v_2,$	$\Phi(\{\beta\}) = v_1, \Phi(\{\alpha, \beta\}) = v_4,$ $\Phi(\{\gamma\}) = v_1, \Phi(\{\beta, \gamma\}) = v_3,$ $\Phi(\{\alpha\}) = v_2, \Phi(\{\alpha, \gamma\}) = v_4,$ $\Phi(\emptyset) = v_4, \Phi(\{\alpha, \beta, \gamma\}) = v_4$
defeats	$-$	$defeat_{VAF}(\beta, \alpha)$ $defeat_{VAF}(\gamma, \alpha)$	$\mathbf{def}(\{\beta\}, \{\alpha\}),$ $\mathbf{def}(\{\gamma\}, \{\alpha\}),$ $\mathbf{def}(\{\alpha\}, \{\beta, \gamma\})$
maxi-defeats	$-$	$-$	$\mathbf{def_M}(\{\alpha\}, \{\beta, \gamma\})$
admissible	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$	$\emptyset, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$	$\emptyset, \{\alpha\}$
accrued	$-$	$-$	$\{\alpha\}$
preferred	$\{\beta, \gamma\}$	$\{\beta, \gamma\}$	$\{\alpha\}$

Note that Φ_{VAF}, $defeat_{VAF}$ are the valuation function and defeat relation as defined in VAF [22].

Table 3.6: Summary of Example 3.3.1

and "Bill has robbed someone, therefore he should be jailed." respectively. The attacks rela-
tionship capturing the interplay between the arguments are represented as $\{attacks(\alpha,\beta),$
$attacks(\beta,\alpha), attacks(\alpha,\gamma)$ and $attacks(\gamma,\alpha)\}$. Given the set of arguments and attack re-
lation, the conflict-free set of arguments for AF, VAF and $PAAF$ are $\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta, \gamma\}$.
In other words, the set of conflict-free sets of arguments is the same for all three frameworks.

Let us now assume the existence of a set of abstract values, a total ordering on the set and
a valuation function Φ as in the previous example. As played out in the motivating exam-
ple 1.2.2, we know that Bill is most likely to go to jail due to the crimes that he has committed.
In such a situation, the accrual of arguments strengthen the accrued set of arguments. To
capture the preference on the arguments, we will assign v_1 to $\{\beta, \gamma\}$, v_2 to $\{\alpha\}$, v_3 to both
$\{\gamma\}$ and $\{\beta\}$ and v_4 to all other combinations. Table 3.7 presents a summary of the outcomes
for each framework. Note that the preferred extension for AF and $PAAF$ is $\{\beta, \gamma\}$ while
$\{\alpha\}$ is the preferred extension for VAF. This illustrates that VAF is not able to capture
the intention of the illustrating example while AF selected the maximal admissible set which
coincidentally happened to be the appropriate answer.

The two examples illustrate several fundamental differences between AF, VAF and $PAAF$.
The fundamental difference between AF and $PAAF$ is the use of preferences in $PAAF$.
This difference allows for the ordering of arguments. The fundamental difference between
VAF and $PAAF$ is how these preferences are utilised. In $PAAF$, the preference values are

	AF	VAF	$PAAF$
AR	$\{\alpha, \beta\,\gamma\}$		
$attacks$	$\{attack(\alpha,\beta), attack(\beta,\alpha), attack(\alpha,\gamma), attack(\gamma,\alpha)\}$		
$conflict\text{-}free$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta,\gamma\}$		
V	$-$	$\{v_1, v_2, v_3, v_4\}$	
$\leq$	$-$	$\{v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4\}$	
Φ	$-$	$\Phi_{VAF}(\alpha) = v_2,$ $\Phi_{VAF}(\beta) = v_3,$ $\Phi_{VAF}(\gamma) = v_3,$	$\Phi(\{\alpha\}) = v_2, \Phi(\{\beta,\gamma\}) = v_1,$ $\Phi(\{\beta\}) = v_3, \Phi(\{\alpha,\beta\}) = v_4,$ $\Phi(\{\gamma\}) = v_3, \Phi(\{\alpha,\gamma\}) = v_4,$ $\Phi(\emptyset) = v_4, \Phi(\{\alpha,\beta,\gamma\}) = v_4$
$defeats$	$-$	$defeat_{VAF}(\alpha,\beta)$ $defeat_{VAF}(\alpha,\gamma)$	$\mathbf{def}(\{\alpha\},\{\beta\}),$ $\mathbf{def}(\{\alpha\},\{\gamma\}),$ $\mathbf{def}(\{\beta,\gamma\},\{\alpha\})$
$maxi\text{-}defeats$	$-$	$-$	$\mathbf{def_M}(\{\beta,\gamma\},\{\alpha\})$
$admissible$	$\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\beta,\gamma\}$	$\emptyset, \{\alpha\}$	$\emptyset, \{\beta,\gamma\}$
$accrued$	$-$	$-$	$\{\beta,\gamma\}$
$preferred$	$\{\beta,\gamma\}$	$\{\alpha\}$	$\{\beta,\gamma\}$

Note that Φ_{VAF}, $defeat_{VAF}$ are the valuation function and defeat relation as defined in VAF [22].

Table 3.7: Summary of Example 3.3.2

utilised in construction of a different definition for defeat and the introduction of a notion of maximal defeater. Furthermore, the preference values are utilised to ordering the sets of arguments (extensions). Such an approach allows $PAAF$ to capture situations where the accrual of arguments yields a stronger position such as example 1.2.2 and more importantly, situations where the accrual of arguments yields a weaker position such as example 3.1.1.

3.3.2 Sources, Credibility And Trust

The aim of sources in argument is to determine the origin of the argument. The ability to distinguish between different sources is an important requirement in argumentation as the acceptability of an argument is directly related to the credibility source of the argument (highlighted in motivation example 1.2.3 and 1.2.4). The issue with many argumentation systems and frameworks is the inability to distinguish between different sources of arguments. This results in arguments being treated with the same level of credibility and trust. Furthermore, in situations where arguments are identical bar the source of the argument, the inability to differentiate between argument sources results in what seems like duplication of arguments and in some systems, repetition of arguments are explicitly disallowed.

In $PAAF$, the two extraction functions (Θ_a, Θ_u) in conjunction with the defeats relation partially address this problem by only considering interacting arguments. The use of the

preference values as informed representation of argument sources will further address the credibility issue. We will describe how this can be addressed with $PAAF$. More details on how this is achieved can be found in chapter 5.2. Let us consider the two motivation example 1.2.3 and 1.2.4.

Let us first consider example 1.2.3 and the issues highlighted in the example. Recall that arguments were presented by three individuals Tom, Dick and Harry regarding whether Bill should be jailed for his crimes. Each of these individuals has different degree of trustworthiness. The different degree of trust will influence the outcome of the debate. Let us view the abstract values as unique identifier representing the collection of individuals. We then associate each unique identifier with the argument. We can then order the abstract values such that the credible participants are more preferred. As such, the result will be that arguments from the more preferred individual will be more preferred or deemed more trustworthy.

In example 1.2.4, the basic issue highlighted in the example is multiple repeated identical arguments from different sources. Again, let us view the abstract values as unique identifier representing the collection of individuals. By associating each unique identifier with an argument, we are now able to uniquely identify each of the repeating. We can again order the abstract values such that the more credible participants are more preferred. Utilising such an approach, captures an ownership relation between the source and the argument.

Finally, treating the abstract values as informed representation of argument sources allows us to capture context sensitive creditability within an argumentation framework. This can be achieved by using a range of context specific valuation functions and ordering on the abstract values while retaining the underlying argumentation machinery. This highlights the flexibility of the $PAAF$ framework and its applicability in addressing a range of real-world problems.

In following section, we will present a summary of the main concepts discussed in this chapter.

3.4 Summary

In this chapter, we presented the abstract preference-based accrual argumentation framework ($PAAF$). Within this chapter, we have motivated the use of preference values in the accrual of arguments. This framework utilises abstract values (preference values) as additional reasoning capabilities when a cycle is encountered between attacking arguments. This

is achieved by the introduction of the defeats and maximally defeats relations. The maximally defeats relation is of particular interest as it captures situations where the accrued set of arguments strengthen a conclusion as well as situations where the accrued set of arguments weakens a conclusion. The framework also utilises the preference values to order the set of admissible sets of arguments and the set of preferred extensions.

Furthermore, we performed a discussion on the result of adding the injective or surjective constraints on the valuation function. With the inclusion of these additional constraints, we highlighted the ability for the framework to generate unique preferred extensions as well as multiple preferred extension. We also presented the required conditions for the generation of a unique preferred extensions and multiple preferred extension. Hence, showing that $PAAF$ is applicable to a range of problems.

Using examples, a comparison between the abstract argumentation framework (AF), the value-based argumentation framework (VAF) and the abstract preference-based accrual argumentation framework ($PAAF$) is presented.

Finally, we presented a discussion on the use of preferences values as informed representation of argument source, credibility and trust.

The next chapter will present the theoretical work on mixed-initiative argumentation.

4

Mixed-Initiative Argumentation Framework[1]

"Rhetoric is nothing,
but reason well dressed and argument put in order."
– Jan Zamoiski

4.1   Introduction

In the previous chapter, we introduced the preference-based accrual abstract argumentation framework ($PAAF$). In this chapter, we will focus on the details of a mixed-initiative argumentation framework ($MIAF$). The use of argumentation for decision support is not new, with a long history of studies such as [12, 14–16, 21, 60, 81, 128]. The use of argumentation

---

[1]Some work presented in this chapter also appeared in [37]

for decision support suffers from two key problems. Firstly, the background knowledge required (for instance, to determine inconsistency, or attack relations, between arguments) is often hard to come by and needs to be manually encoded (often an expensive proposition). Secondly, the bases for decision making often end up being inconsistent over a series of decisions (e.g. arguments X is preferred to argument Y in obtaining a given decision, but Y is preferred to X in obtaining the next). There is a clear need for a formal framework for the management of justifications (or rationale) in decision support to maintain a certain level of quality of the resulting decisions. The quality assurance of decisions entails the retention of decision justification such that retrospective analysis can be performed on decisions to assess the "correctness" of these decisions. The use of decision justification for such retrospective analysis maintains consistency between past and current decisions hence maintaining the quality of the background knowledge for future decisions comparison. The main focus of this chapter is to propose a *mixed-initiative argumentation* approach that address these issues.

General focus of mixed-initiative interaction research centres around the development of methods that enable computing systems to support an efficient, natural interleaving of contributions between participating parties with the aim of converging onto solutions to problems. Mixed-initiative interaction is described as a flexible interaction strategy in which parties (human or computer) take initiatives in contributing their best suited-skill set at the most appropriate time to solving a problem, achieving a goal or coming to a joint understanding [4,44,74]. In other words, such a system would carry out problem solving tasks on behalf of user where both the user and the system can take the initiative by directing the problem solving task. Allen [4] presented several different levels of mixed-initiative interactions.

Our mixed-initiative argumentation framework falls into the "sub-dialogue initiation" level of interaction. The sub-dialogue initiation level of interaction classifies agents or systems with the ability to initiate sub-dialogues for clarification and corrections. At this level of interaction, the system is able to initiate a sub-dialogue for the purpose of clarification or resolving a problem. In essence, providing the ability for a system to interrupt a current process for the purpose of problem clarification or resolution. Such dialogue may take several interactions to complete. Hence, the system has temporarily taken the initiative until the problem has been resolved. As the name suggests, the "initiative" in the problem solving process can come in equal measure from the "system" and the "user". Such a framework permits the interleaving of *decision generation* with *decision justification*. Decision generation steps involve classical argumentation, where the "winning" argument(s) are identified by the argumentation machinery. Decision justification steps are more complicated, and require an "inversion" of the machinery for decision generation. In the decision justification process,

the user selects the "winning" arguments, and is then prompted to *justify* the decision by updating the background knowledge that is brought to bear in decision generation. The steps involved in the decision justification process are as follows:

1. The user selects or provides the "winning" arguments. This set of "wining" arguments is the support for particular class of decisions. We view the decisions within this class as equivalent and indistinguishable hence essentially, the set of arguments constitutes a decision. This set of arguments is compared with that obtained from the argumentation machinery for decision generation process.

2. If the two sets of arguments differs, the user is prompted to *justify* his/her decision. These justifications can be expressed as relations between arguments such as attack, defeat or preference relations.

3. The new justifications are then added to the argumentation theory and a new set of "winning" arguments is generated by the argumentation machinery and a comparison is performed again.

4. These steps are repeated until the argumentation machinery is able to generate as a unique outcome, the set of arguments identified by the user as the "winning" arguments.

Furthermore, as the knowledge encoded in the argumentation theory is hard to come by, we endeavour to retain as much of this knowledge as consistently possible. To achieve this, we draw inspiration from the principle of *minimal change*, that has a long history in philosophy and in AI approaches to theory change, to underpin this process. In devising the machinery for decision justification, we are interested in answering the following question: how might we minimally modify the argumentation system (specifically the background knowledge) in order to obtain one that would generate the (user-specified) "winning" argument(s) when used in the decision-generation mode?

In section 4.1.1, we will motivate our approach by using an example from clinical group decision making. In section 4.2, we will introduce the formal details of the mixed-initiative argumentation framework as well as a set of minimal change procedures for modifying the argumentation theory. Section 4.4 presents several key concepts. Firstly, a set of properties for mixed-initiative argumentation are presented. Secondly, we present a discussion on the flexibility of the mixed-initiative argumentation framework. In section 4.5, we present a summary of key ideas and concepts presented in this chapter.

4.1.1 Illustrating Example

In this example, we will demonstrate the need for a formal framework for the management of justifications (or rationale) in decision support. Furthermore, we illustrate that the corpus of knowledge utilised for decision making in the "real-world" is a state of constant flux. Hence over a sequence of decisions, it is evident that an update or revision is required on the background knowledge that is brought to bear by any decision support system in generating a decision. Let us consider an extract from a medical group decision session (Figure 4.1). The discussion is on a patient with localised breast cancer. The discussion involves several medical specialists (MD_1, MD_2, MD_3) debating on the best treatment for the disease. Decision-making in a multidisciplinary team (MDT) reflects a process of debate where "competing" clinicians argue for particular management strategies. While this process should bring to bear the pertinent evidence from the published medical literature and the personal results of clinicians, it can also include hearsay, personal preference and prejudice.

Consider the group decision exchange as a sequence of three interactions pointing to a decision at each interaction. Firstly, let use focus on the issue of how new evidence can support alternate viewpoints. Let us assume that initially, the background knowledge is empty. Let us first consider the arguments A_1, A_3, A_4. After the first two exchanges, the recommendation is to always perform a lumpectomy. Furthermore, the recommendation also argue against the need for radiotherapy, hence MD_1 has decided that no additional treatment was required for the patient. However, when presented with the additional evidence (A_7) that proactive treatment will reduce the 10-year relapse rate, MD_1 assimilated this new evidence and consequently altered the recommendation. Hence, the new recommendation is to perform lumpectomy and then follow by chemotherapy.

Secondly, the use of previous justification to constrain later decisions such that the sequences of decision are consistent. The two justifications provided by MD_1 for not performing radiotherapy or chemotherapy consistently point to the lack of improvement in the overall death rate. In this instance, these two justifications are consistent. Thirdly, argumentation is not a single facet process. The interaction shows that the MDs have argued over treatment by considering the presence and timing of local, regional and distant disease recurrence, as well as death. By emphasising and managing the justification of decisions from explicit evidence, we have constrained all future decision making to match this consistency. The use of argumentation ensures that the sequences of decisions are justified in a consistent evidence-based manner. Since evidence accumulates with time, any system must include the ability to subsume new knowledge, to incorporate the local introduction or loss of techniques or experience, and to interface with clinicians regarding their decision-making. The system should

Disease Definition: Breast Cancer
Localised

	Question: Should surgery be performed? Answer: Decision is to perform lumpectomy
MD_1	: (A_1) Given the type of cancer, I believe that we should undertaken a lumpectomy.
MD_2	: (A_2) Agree.

	Question: Is radiotherapy required? Answer: Decision is to do nothing
MD_1	: (A_3) I have undertaken a lumpectomy with clear margins and an auxiliary lymph node dissection, with no positive nodes, therefore I believe that there is no need of other treatment
MD_1	: (A_4) No radiotherapy required. Given her age, tumour size, grade, margin status, the local recurrence risk is 1.3% with radiotherapy and 4.5% without radiotherapy, and no benefit in overall survival. The local control benefit is not large enough to justify the offer of treatment
MD_2	: (A_5) Agree.

	Question: Is chemotherapy required? Answer: Decision is to perform chemotherapy
MD_1	: (A_6) No chemotherapy required, her 10 year death rate will only reduce from 5% to 4% with chemotherapy and has only a small effect on the local recurrence rate (now 3.2%)
MD_3	: (A_7) I disagree. I think she should have chemotherapy, because while the death rate is similar, her 10-year relapse rate will fall from 25% to 11% with chemotherapy, meaning there is more life without cancer.
MD_1	: (A_8) Yes. Good point, I had not considered the relapse numbers. She should have chemotherapy.

	Given another patient with the exact disease: Decision is to perform lumpectomy follow with chemotherapy
MD_1	: (A_9) Given the type of cancer, I believe that we should undertaken a lumpectomy.
MD_1	: (A_{10}) I have undertaken a lumpectomy with clear margins and an auxiliary lymph node dissection, with no positive nodes. However, I think she should have chemotherapy, because while the death rate is similar, her 10-year relapse rate will fall from 25% to 11% with chemotherapy, meaning there is more life without cancer.

Figure 4.1: Clinical Decision Support: Breast Cancer

detect where its decision is at odds with the clinician's and assimilate evidence provided for the discrepancy. This illustrates the need for argumentation systems to be open to new facts, changing rules and preferences.

The example also illustrates that an argumentation system should evolve over time, accumulating past decision as justification for future decisions. However, it is also clear that in some instances, we wish to overrule past precedent. In most argumentation and decision support systems presented in the literature, the systems are relatively static. Most systems are open to new facts, however, have difficulties handling changing rules and preferences.

Furthermore, let us now assume that the patient's physician decided to perform radiotherapy. He/She will now have to justify the decision. If we assume that the above discussion did not occur (i.e. empty knowledge base), then the physician only requires to presents arguments for the decision to only perform lumpectomy. However, if the knowledge base consists of the arguments, attack relations and preferences captured from the discussion, then the physician will be required to not only present arguments for the decision to perform lumpectomy but also address all attacks on his/her decision. One can view the sequence of interaction captured in the discussion as "decision generation" mode, if all the arguments, attack relation and preferences exists in the knowledge base (in other words, the knowledge base is complete) and we requested the argumentation system to present us with a decision. Alternatively, if the knowledge base is incomplete, erroneous or an undesired decision generated (due to unforeseen interaction between rules), a decision can be introduced and modification performed on the knowledge base in the "decision justification" mode.

In this spirit, we will discuss how a mixed-initiative abstract argumentation framework can be used to model such discourse. In the next section, we will present an abstract mixed-initiative argumentation framework.

4.2 Formal Framework

In this section, we will present the formal details for the mixed-initiative argumentation framework ($MIAF$). A mixed-initiative argumentation framework consists of three distinct components: argumentation theory, decision generation engine and decision justification engine. The argumentation theory and the decision generation engine exist in any traditional argumentation framework and system. The unique component of this framework is the existence of a decision justification engine. User interaction with the system over a period of

time can be viewed as an interaction sequence $\langle i_1, i_2, i_3, \ldots, i_n \rangle$ consisting of interactions of two types:

1. Decision generation

2. Decision justification

Decision generation involves the generation of extensions. We maybe interested in generating all extensions (with which we might perform sceptical or credulous reasoning). Alternatively, we might only be interested in the preferred extensions. Traditionally, an argumentation system can be considered as a function that takes in several components and generates as its output, a set of extensions supporting different and usually competing decisions. These extensions constitute justifications for decisions. Generally, a particular extension is then selected (external to the argumentation machinery) from this set of extensions by the reasoning agent as justification to the committed decision. In our framework, we denoted this process as decision generation. In figure 4.2, we present a conceptual view of the argumentation system. The inputs can be grouped into three categories: a set of arguments, a conflict theory and a preference theory.
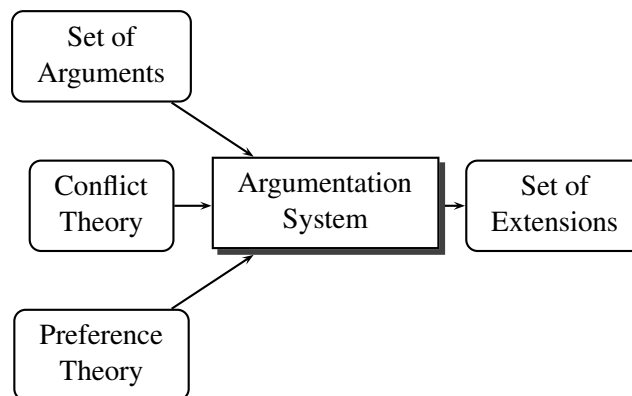

Figure 4.2: Traditional Argumentation System

The output of an argumentation system is sets of arguments that supports possible alternate outcomes. This set of arguments are the set of extensions. We view supports for outcomes as decisions hence we will use the terms extensions and decisions interchangeably. Similar to studies ([25, 52, 77, 78]) into abstract argumentation, we will not assume any knowledge regarding the structure of arguments. An argumentation theory is defined as:

Definition 4.2.1. *An **argumentation theory** is a triple $\langle AR, Conf, Pref \rangle$ where:*

- *AR is a set of arguments $\{\alpha_1, \ldots, \alpha_i, \ldots, \alpha_n\}$*

- $Conf$ *is the conflict theory consisting of non-transitive relations/assertions of the form* $conf(\alpha_i, \alpha_j)$ *where* $\alpha_i, \alpha_j \in AR$

- $Pref$ *is the preference theory consisting of non-transitive relations/assertions of the form* $pref(A_i, A_j)$ *where* $A_i, A_j \subseteq AR$

Given $A_i, A_j \subseteq AR$, we will denote $pref(A_i, A_j)$ to mean A_i is preferred to A_j. The assertion $conf(\alpha_i, \alpha_j)$ denotes that argument α_i conflicts with argument α_j. Note that given a language L with an associated entailment relation $\models_L$, any theory T composed of sentences in L induces a consistency theory. Given such a consistency theory, assertions in $Conf$ can be computed. In the absence of a background theory, such sentences need to be explicitly (user-) defined. Preference assertions are always explicitly user-defined. Since we are not viewing assertions in AR as a truth functional language, we need to explicitly perform the assertions. The requirement for non-transitivity in the conflict theory is important as conflict are not transitive. Note that the requirement for non-transitivity in the preference theory allows for alternate solutions to address the accrual of arguments as highlighted in chapter 3 preference theory however, transitivity can be achieved by additional assertions into the preference theory.

Definition 4.2.2. *Given an* $AT = \langle AR, Conf, Pref \rangle$. $Deft$ *is a set of non-transitive relations/assertions of binary of the form* $deft(\alpha_i, \alpha_j)$ *such that for all* $\alpha_i, \alpha_j \in AR$, $deft(\alpha_i, \alpha_j) \in Deft$ *if and only if* $(conf(\alpha_i, \alpha_j) \in Conf$ *and* $conf(\alpha_j, \alpha_i) \notin Conf)$ *or* $(conf(\alpha_i, \alpha_j) \in Conf$ *and* $conf(\alpha_j, \alpha_i) \in Conf$ *and* $pref(\{\alpha_j\}, \{\alpha_i\}) \notin Pref)$

Note that the use of conflict in this situation is not symmetrical because it encompasses different forms of attack that might be directional. More detail of this discussion can be found in chapter 5.2.2. The intuition behind defeat is to identify successful conflicts. As such, defeat captures situation where conflict that are directional and in the situations where conflict is symmetrical, the preferences theory is utilised to determine the outcome. This definition of defeat differs from those defined in [22, 49, 50, 78]. The primary difference is due to the usage of preferences when defining the notion of defeat. In this setting, our aim is not to dilute the underlying machinery of the argumentation framework. The use of preferences is therefore used sparsely, only when the underlying attack relation is not able to determine a "winner", should the preferences theory fire to determine the outcome. Note also that there are no requirements for the preferences to be totally ordered. This allows for conflict to be left unresolved or ignored and hence a winning argument to be undecided. In some circumstance, the ability to ignore conflict or deferring making judgement until later may yield a better final result.

Definition 4.2.3. *Given an $AT = \langle AR, Conf, Pref \rangle$ and a set of arguments $S \subseteq AR$, S is an **extension** if and only if it satisfies the following:*

- *(Absences of conflict) There does not exist any argument $\alpha_i, \alpha_j \in S$ such that $deft(\alpha_i, \alpha_j) \in Deft$.*

- *(Admissible) For all arguments $\alpha_i \in S$, if there exist an argument $\alpha_j \in AR$ and $deft(\alpha_j, \alpha_i) \in Deft$ then there exist an argument $\alpha_k \in S$ such that $deft(\alpha_k, \alpha_j) \in Deft$.*

- *(Maximality) There is no set S' that satisfies first two conditions (absences of conflict, admissible) such that $S \subset S' \subseteq AR$ and $pref(S', S) \in Pref$.*

In the following section, we will use Ext_{AT} to denote the set of all extensions of an argumentation theory AT.

Note that the preferred extension defined in [52] is a special case of the extension definition given above. They coincide if the preference theory is empty. Given a finite set of distinct decision options O and a function $D : 2^{AR} \longrightarrow O$ linking sets of arguments to the decision options, each extension is mapped to a decision. Since each extension is maximal, sub-maximal sets of arguments are ultimately defeated or superseded by some extension. We can view that the decisions associated with each extension as the only supported choices. Hence for this discussion, we will use the terms extension and decision interchangeably.

A mixed-initiative argumentation system can be viewed as a revision function which takes as input; the original argumentation theory, a pre-selected extension, any new arguments, new conflict theory or new preference theory and outputs a revised argumentation theory with the pre-selected extension as its output. We denoted this process as decision justification. In figure 4.3, we present the conceptual view of a mixed-initiative argumentation system.

Definition 4.2.4. *Given $\mathcal{AT}$, the class of all argumentation theories, and $\mathcal{AR}$, the universe of arguments, a **mixed-initiative argumentation system** is defined as $\langle A_{gen}, A_{just} \rangle$ where:*

1. *A_{gen} is a **decision generation function** such that $A_{gen} : \mathcal{AT} \longrightarrow 2^{\mathcal{AR}}$. Intuitively, A_{gen} selects the "winning" extension, given that an argumentation theory may in general support multiple extensions.*

2. *A_{just} is a **decision justification function** such that $A_{just} : \mathcal{AT} \times 2^{\mathcal{AR}} \longrightarrow \mathcal{AT}$. Intuitively, A_{just} takes an argument theory AT and a user-specified set of arguments,*
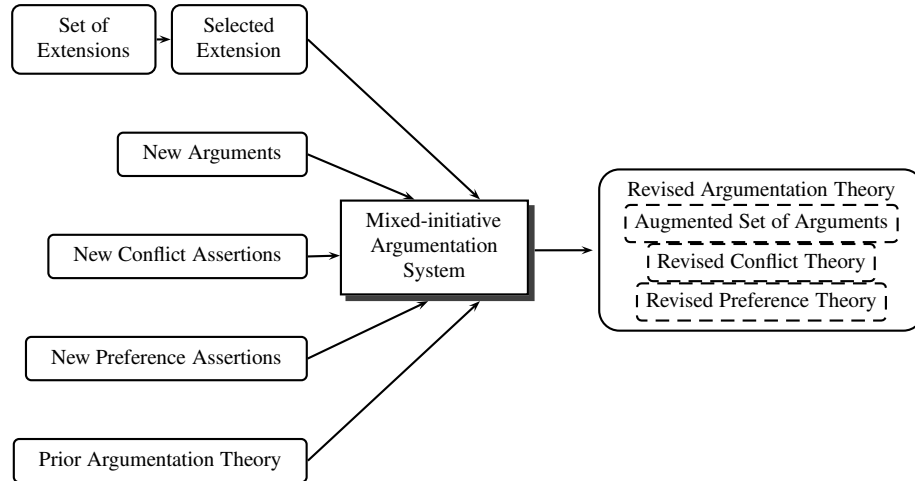
Figure 4.3: Mixed-Initiative Argumentation System in decision justification mode

> *and generates a revised argumentation theory AT', such that the input set of arguments is the "winning" extension if A_{gen} were to be applied to AT'*

We use the term "winning" extension in the definition above (as opposed to "preferred" extension, for instance) mainly because our definition of an extension already incorporates the application of the preference theory. We admit the possibility of multiple extensions, hence the identification of a unique "winning" extension must involve the application of criteria (such as user choice) extraneous to those encoded in an argumentation theory. The main objective of a mixed-initiative argumentation system is to perform group decision support activities. A_{just} is used to construct the rationale for supporting a selected decision. The user-specified extension is externally provided. Since arguments are linked to the decision options by the function D, we will refer to the *user-specified extension* as a *user-specified decision*.

Definition 4.2.5. *Given $AT = \langle AR, Conf, Pref \rangle$ and a (user-specified) decision S, AR is S-**complete** if and only if $S \subseteq AR$.*

Definition 4.2.6. *A decision justification function A_{just} is a **preference modifying** decision justification function if and only if for every argumentation theory $AT = \langle AR, Conf, Pref \rangle$ and every decision S, if S is not the unique decision of AT, $A_{just}(AT, S) = \langle AR, Conf, Pref' \rangle$ where $Pref \neq Pref'$, provided that AR is S-complete.*

Definition 4.2.7. *A decision justification function A_{just} is a **conflict modifying** decision justification function if and only if for every argumentation theory $AT = \langle AR, Conf, Pref \rangle$ and every decision S, if S is not the unique decision of AT, $A_{just}(AT, S) = \langle AR, Conf', Pref \rangle$ where $Conf \neq Conf'$, provided that AR is S-complete.*

In following section, we will firstly describe the decision justification procedure. The described decision justification procedure is then decomposed into more detail sub procedures. For each procedure, we will illustrate a set of interactions between the conflict theory and preference theory as well as providing a range of revision solutions to the different cases. The procedures are also presented more formally in Appendix A. Finally, we will present the properties of the procedures.

4.3 Decision Justification Procedures

In a mixed-initiative argumentation system, user interaction with the system over a period of time can be viewed as an interaction sequence $\langle i_1, i_2, i_3, \ldots, i_n \rangle$ consisting of interleaving interactions of two types: decision generation process and decision justification steps. The general procedure is illustrated in figure 4.4.
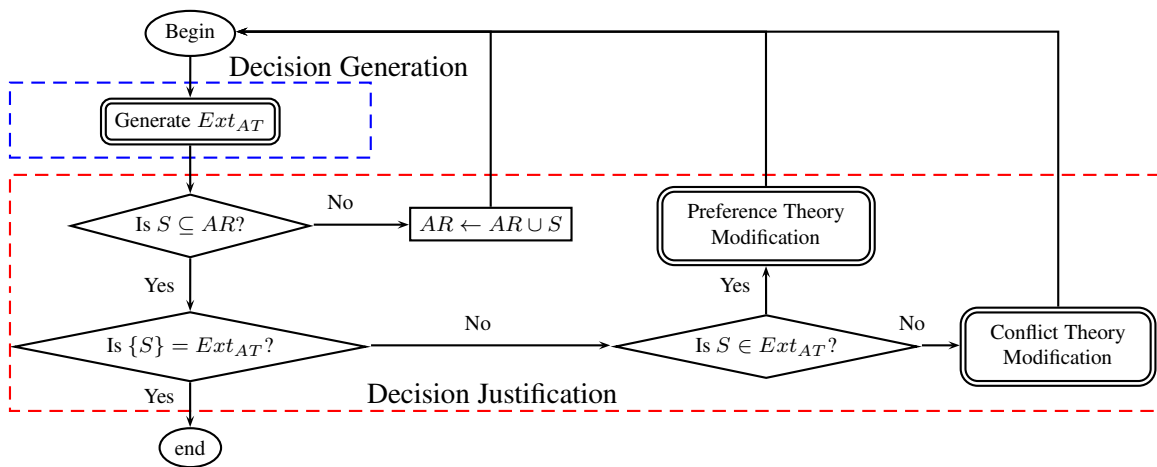

Figure 4.4: Mixed-Initiative Argumentation System Procedure

We note that the definition for an extension above directly provides a decision generation procedure. A range of approaches can be utilised to optimising such a procedure, but we do not consider these here due to space restrictions (however, some of these have been utilised in the implementation of the tool described in chapter 5.3.2). For instances, the procedure of [51, 139] could be extended to account for a preference theory to suit our requirements.

4.3.1   Preference Modifying Decision Justification Procedure

Firstly, let us focus on the modification of the preference theory. The aim of the preference modifying decision justification procedure is to rearrange the preference ordering such that user specified decision S is the most preferred. The preference modifying decision justification procedure can be informally described as follows: For every decision $A_i \in (Ext_{AT} \setminus \{S\})$, if S is not preferred to A_i, then insert the assertion $pref(S, A_i)$ into $Pref'$ and remove the assertion $pref(A_i, S)$ from $Pref'$. Finally, we return the modified $Pref'$. Since $Pref$ consists of non-transitive binary assertions, we need not worry about the consequence of inverting an assertion. A simple example illustrating the results of such insertion is shown in example 4.3.1. A more formal description of the procedure can be found in procedure A.2 (in the appendix).

Example 4.3.1. *Given* $AT = \langle AR, Conf, Pref \rangle$, *assume that* $Ext_{AT} = \{A_1, A_2, S\}$ *and* $Pref = \{pref(A_1, S), pref(A_1, A_2)\}$. *Removing the assertion* $pref(A_1, S)$ *from* $Pref'$ *and subsequently adding the assertion* $pref(S, A_1)$ *and* $pref(S, A_2)$ *to* $Pref'$ *results in the modified* $Pref' = \{pref(S, A_1), pref(S, A_2), pref(A_1, A_2)\}$. *This will result in an ordering where* S *is the most preferred hence* S *is the unique extension of* $AT = \langle AR, Conf, Pref' \rangle$

Properties of The Preference modifying Decision Justification Procedure

We will now consider the termination, soundness, completeness and minimal change properties of the preference modifying decision justification procedure. Firstly, let us consider the termination properties of the preference modifying decision justification procedure. The termination property illustrates that the preference modifying justification procedure will eventually halt.

Theorem 4.3.1 (Termination). *Given an argumentation theory* $AT = \langle AR, Conf, Pref \rangle$, *if* AR *is a finite set then the preference modifying decision justification procedure terminates.*
Proof:
The preference modifying decision justification procedure terminates if all loops terminates. Given the procedure evaluates each $A_i \in (Ext_{AT} \setminus \{S\})$, the preference modifying decision justification procedure terminates when all $A_i \in (Ext_{AT} \setminus \{S\})$ has been considered. Let us consider two cases:

Case 1: Let us assume that Ext_{AT} is finite and the procedure does not terminate. Given that Ext_{AT} is a finite set, then eventually all $A_i \in (Ext_{AT} \setminus \{S\})$ will be evaluated. Hence, the procedure terminates. Thus violating the assumption.

Case 2: Let us assume that Ext_{AT} is not finite and the procedure terminates. Given that Ext_{AT} is an infinite set, then the procedure will continue evaluating elements $A_i \in (Ext_{AT} \setminus \{S\})$ and never terminate. Thus violating the assumption. □

Let us now consider the soundness property of the preference modifying decision justification procedure. The soundness property illustrates that the procedure will always return a modified preferences theory such that the user-specified decision is the unique decision.

Theorem 4.3.2 (Soundness). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S and S is not the unique decision of AT, the preference modifying decision justification procedure is sound.*

Proof:

The preference modifying decision justification procedure is sound if and only if the preference modifying decision justification procedure realises a preference modifying decision justification function. In other words, the preference modifying decision justification procedure returns a $Pref'$ such that S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$. S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$ if and only if it satisfies the following conditions:

1. $S \in Ext_{AT'}$: Let us assume that $S \in Ext_{AT}$ and $S \notin Ext_{AT'}$. Let us consider two cases:

Case 1: There exists $A_i \in (Ext_{AT} \setminus \{S\})$ s.t $pref(A_i, S) \in Pref'$. However, the preference modifying decision justification procedure inserts the assertion $pref(S, A_i)$ into $Pref'$ for all $A_i \in (Ext_{AT} \setminus \{S\})$ and removes all the assertions $pref(A_i, S)$ from $Pref'$ whenever S is not preferred to A_i. Hence there does not exist $A_i \in (Ext_{AT} \setminus \{S\})$ s.t $pref(A_i, S) \in Pref'$. Thus, the assumption is violated.

Case 2: There exists $A_i \in (Ext_{AT} \setminus \{S\})$ s.t $pref(S, A_i) \in Pref'$ and $pref(A_i, S) \in Pref'$. Since the preference modifying decision justification procedure inserts the assertion $pref(S, A_i)$ into $Pref'$ for all $A_i \in (Ext_{AT} \setminus \{S\})$ and remove all the assertions $pref(A_i, S)$ from $Pref'$ whenever S is not preferred to A_i, we can conclude that if $pref(S, A_i) \in Pref'$ and $pref(A_i, S) \in Pref'$ then $pref(S, A_i) \in Pref$ and $pref(A_i, S) \in Pref$. From the maximality requirement in definition 4.2.3, we can conclude that S is not an extension. This entails that $S \notin Ext_{AT}$. Thus, violating the assumption.

2. For all $A_i \in (Ext_{AT} \setminus \{S\}), pref(S, A_i) \in Pref'$: Let us assume that there exists $A_i \in Ext_{AT}$ s.t. $pref(S, A_i) \notin Pref'$. This entails that S is not preferred to A_i. However, the preference modifying decision justification procedure inserts the assertion $pref(S, A_i)$ into $Pref'$ for all $A_i \in (Ext_{AT} \setminus \{S\})$ and removes all the assertions $pref(A_i, S)$ from $Pref'$ whenever S is not preferred to A_i. Thus, the assumption is violated. $\square$

Let us now consider the existence property of the preference modifying decision justification function. The existence property illustrates that there exists at least one preference modifying decision justification function that will be able to perform the required modification such that the user-specified decision is the unique decision of the modified argumentation theory.

Theorem 4.3.3 (Existence). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S, if S is not the unique decision of AT, then there exists at least one preference modifying decision justification function such that S is the unique decision of $A_{just}(\langle AR, Conf, Pref \rangle, S)$.*

Proof:
Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S, if S is not the unique decision of AT, then there exists $AT' = \langle AR, Conf, Pref' \rangle$ such that S is the unique decision. Since there exists $AT' = \langle AR, Conf, Pref' \rangle$ such that S is the unique decision then there exists a preference modifying decision justification function such that S is the unique decision of $A_{just}(\langle AR, Conf, Pref \rangle, S)$. $\square$

We will now consider the minimally modification property of the preference modifying decision justification function. Given a modified preference theory $Pref'$ derived from $Pref$, we consider $Pref'$ to be a minimally modified preference theory if there does not exist any subset of $Pref'$ that will yield the same unique user-specified decision.

Definition 4.3.1. *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$ and a (user-specified) decision $S \subseteq AR$, a preference theory $Pref'$ is a minimally modified preference theory with respect to AT and S if and only if:*

1. *S is the unique decision of $\langle AR, Conf, Pref' \rangle$.*

2. *There exists no $Pref''$ such that $Pref \ominus Pref'' \subset Pref \ominus Pref'$ and S is the unique decision of $\langle AR, Conf, Pref'' \rangle$.*

Note that $\ominus$ denotes symmetric set difference.

Note that a minimal change to a conflict theory can also be expressed with respect to set cardinality by using $|Pref \ominus Pref''| < |Pref \ominus Pref'|$. From the definition above, we can assume that starting with $pref(\{\alpha_i\}, \{\alpha_j\})$, both removing $pref(\{\alpha_i\}, \{\alpha_j\})$ from $Pref$ and replacing $pref(\{\alpha_i\}, \{\alpha_j\})$ with $pref(\{\alpha_j\}, \{\alpha_i\})$ represent the same "quantum of change".

Let us now consider the minimality property of the preference modifying decision justification function. The minimality property illustrates that the preference modifying decision justification procedure will produce a result that is a minimal repair to the original preferences theory such that the user-specified decision is the unique decision.

Theorem 4.3.4 (Minimal). *Given an argumentation theory* $AT = \langle AR, Conf, Pref \rangle$, *a (user-specified) decision S and S is not the unique decision of AT, the preference modifying decision justification procedure generates a preference theory $Pref'$ that is a minimally modified preference theory with respect to AT and S. As well $\langle AR, Conf, Pref' \rangle$ is a minimally (w.r.t set inclusion and set cardinality) modified argumentation theory with respect to AT and S.*

Proof:

$Pref'$ is a minimally (w.r.t set inclusion and set cardinality) modified preference theory with respect to AT and S if and only if it satisfies the following properties:

1. S is the unique decision of $\langle AR, Conf, Pref' \rangle$. Let us assume that S is not the unique decision of $\langle AR, Conf, Pref' \rangle$. This entails that the preference modifying decision justification procedure is unsound and returns a preference theory $Pref'$ such that S is not the unique decision. However, theorem 4.3.2 shows that the preference modifying decision justification procedure is sound and will always return a $Pref'$ such that S is the unique decision of $\langle AR, Conf, Pref' \rangle$. Thus violating the assumption.

2. There exists no $Pref''$ such that $Pref \ominus Pref'' \subset Pref \ominus Pref'$ and S is the unique decision of $AT'' = \langle AR, Conf, Pref'' \rangle$. Let us assume that there exists a $Pref''$ such that $Pref \ominus Pref'' \subset Pref \ominus Pref'$ and S is the unique decision of AT''. This entails that there exists $A_i \in (Ext_{AT} \setminus \{S\})$ s.t. $pref(S, A_i) \in Pref'$ and $pref(S, A_i) \notin Pref''$. It then follows that there exists $A_i \in Ext_{AT''}$ that is as preferred as S. Hence, S is not a unique extension of AT''. Thus violating the assumption.

3. There exists no $Pref''$ such that $|Pref \ominus Pref''| < |Pref \ominus Pref'|$ and S is the unique decision of $AT'' = \langle AR, Conf, Pref'' \rangle$. Let us assume that there exists a $Pref''$ such that $|Pref \ominus Pref''| < |Pref \ominus Pref'|$ and S is the unique decision of AT''. This entails that there exists $A_i \in (Ext_{AT} \setminus \{S\})$ s.t. $pref(S, A_i) \in Pref'$ and

$pref(S, A_i) \notin Pref''$. This entails that there exists $A_i \in Ext_{AT''}$ that is as preferred as S. Hence, S is not a unique extension of AT''. Thus violating the assumption. $\quad\square$

Let us now consider the uniqueness property of the minimally modified preference theory. The uniqueness property illustrates that there exists exactly one minimally modified preference theory such that the user-specified decision is the unique decision of the modified argumentation theory.

Theorem 4.3.5 (Uniqueness). *Given an argumentation theory* $AT = \langle AR, Conf, Pref \rangle$, *a (user-specified) decision* S *and* S *is not the unique decision of* AT, *there is exactly one minimally modified preference theory* $Pref'$ *with respect to* AT *and* S *such that* S *is the unique decision of* $AT' = \langle AR, Conf, Pref' \rangle$.

Proof:
There exists exactly one minimally modified preference theory $Pref'$ with respect to AT and S such that S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$ if and only if there does not exist any other distinct minimally modified preference theory $Pref''$ with respect to AT and S such that S is the unique decision of $AT'' = \langle AR, Conf, Pref'' \rangle$. Let us assume that there exists two distinct minimally modified preference theory $Pref'$, $Pref''$ with respect to AT and S such that S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$ and $AT'' = \langle AR, Conf, Pref'' \rangle$. This entails that there exists $A_i \in (Ext_{AT} \setminus S)$ such that $pref(S, A_i) \in Pref'$ and $pref(S, A_i) \notin Pref''$ or $pref(S, A_i) \in Pref'$ and $pref(A_i, S) \in Pref''$. Let us consider the two cases:

  Case 1: $pref(S, A_i) \in Pref'$ and $pref(S, A_i) \notin Pref''$. This entails that $Pref'' \subset Pref'$. From theorem 4.3.4, it follows that $Pref'$ is not a minimally modified preference theory. Thus violating the assumptions.

  Case 2: $pref(S, A_i) \in Pref'$ and $pref(A_i, S) \in Pref''$. This entails that there exists $A_i \in Ext_{AT''}$ that is as preferred as S. Hence, S is not a unique extension of AT''. Thus violating the assumption. $\quad\square$

Let us now consider the completeness property of the minimally modified preference theory. The completeness property illustrates that preference modifying decision justification procedure will return all modified preference theories leading to S being the unique extension.

Theorem 4.3.6 (Completeness). *Given* $AT = \langle AR, Conf, Pref \rangle$, *a (user-specified) decision* S *and* S *is not the unique decision of* AT, *the preference modifying decision justification procedure is complete.*

Proof:

The preference modifying decision justification procedure is complete if and only if it returns all minimally modified $Pref'$ such that S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$. From theorem 4.3.2, it follows that the preference modifying decision justification procedure is sound. Hence any returned $Pref'$ will lead to the condition that S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$. From theorem 4.3.5, it follows that there exists exactly one $Pref'$ that will satisfy the condition such that S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$. Since preference modifying decision justification procedure is sound and returns all $Pref'$ that satisfy the condition such that S is the unique decision of $AT' = \langle AR, Conf, Pref' \rangle$, it is therefore also complete. $\square$

4.3.2 Conflict Modifying Decision Justification Procedure

In the situation where $S \notin Ext_{AT}$, we will need to consider the modification of the conflict theory. Given $AT = \langle AR, Conf, Pref \rangle$, Ext_{AT} and a decision S, the modification of the conflict theory aims at promoting S to being the only extension in Ext_{AT}. To achieve this, all arguments $\alpha_i \in AR \setminus S$ need to be defeated by some $\alpha_j \in S$. The conflict and preference relationship between each $\alpha_i \in AR \setminus S$ and S maybe characterised as one of several possibilities illustrated in Figure 4.5.

Firstly, we determine if $conf(\alpha_i, \alpha_j) \in Conf$ and $conf(\alpha_j, \alpha_i) \in Conf$. If it is the case that $conf(\alpha_i, \alpha_j) \notin Conf$ and $conf(\alpha_j, \alpha_i) \in Conf$ then there is no modification required. Figure 4.5a illustrates such a situation. If it is the case that $conf(\alpha_i, \alpha_j) \in Conf$ and $conf(\alpha_j, \alpha_i) \in Conf$, then we determine if $pref(\{\alpha_i\}, \{\alpha_j\}) \in Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$. If $pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$, then there is no modification required. Figure 4.5b illustrates such a situation. However, if it is the case that $pref(\{\alpha_i\}, \{\alpha_j\}) \in Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$ or the case that $pref(\{\alpha_i\}, \{\alpha_j\}) \in Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \notin Pref$, then we need to either remove the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$ or select some $\alpha_k \in S$ that is not defeated and add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. Figure 4.5e and 4.5f illustrates such a situation. If it is the case that $conf(\alpha_i, \alpha_j) \in Conf$ and $conf(\alpha_j, \alpha_i) \notin Conf$, then we need to select some $k \in S$ and add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. Furthermore, if $pref(\{\alpha_i\}, \{\alpha_k\}) \in Pref$, we need to remove the assertion $conf(\alpha_i, \alpha_k)$ from $Conf'$. Figure 4.5h, 4.5i and 4.5j illustrates such a situation. Finally, if it is the case that $conf(\alpha_i, \alpha_j) \notin Conf$ and $conf(\alpha_j, \alpha_i) \notin Conf$, then we need to select some $k \in S$ and add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. Figure 4.5c illustrates such a situation. Table 4.1 provides a summary of the conflict theory modification decision matrix.
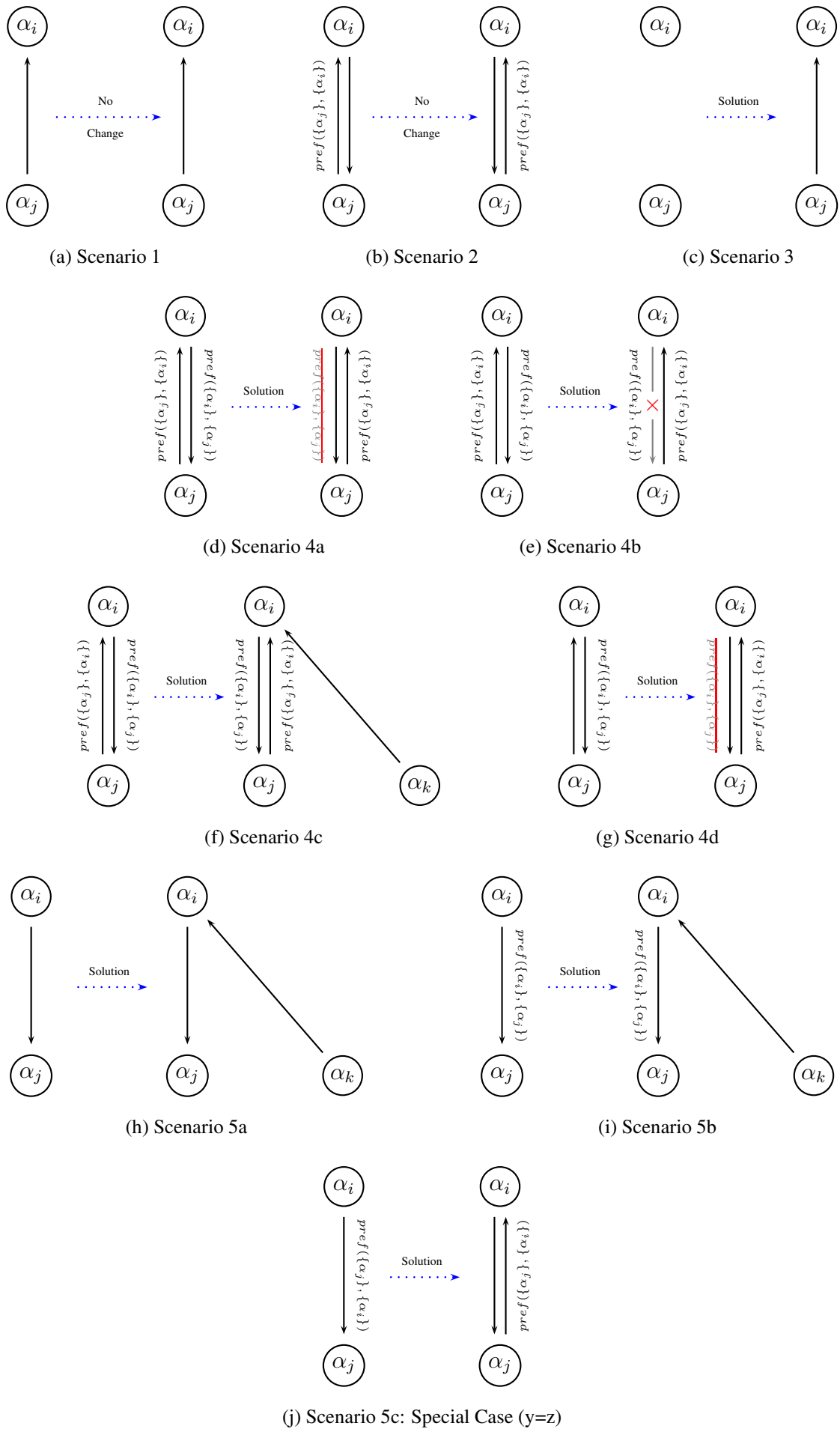
Figure 4.5: Graphical Illustration of Theory Modification Scenarios and Solutions

List of cases to consider for any $x, y \in AR$ and $y \in S$				
Conf		*Pref*		
$conf(\alpha_i, \alpha_j)$	$conf(\alpha_j, \alpha_i)$	$pref(\{\alpha_i\}, \{\alpha_j\})$	$pref(\{\alpha_j\}, \{\alpha_i\})$	Solutions
True	False	True	False	2 or 3
True	False	False	True	2 or 3
True	False	False	False	2 or 3
True	False	True	True	2 or 3
True	True	True	False	2 or 3
True	True	False	True	1
True	True	False	False	2 or 3
True	True	True	True	2 or 3
False	False	-	-	3
False	True	-	-	1

List of solutions
1. Do Nothing
2. Remove the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$ and add the assertion $conf(\alpha_j, \alpha_i)$ into $Conf'$
3. Select some $\alpha_k \in S$ that is not defeated and add the assertion $conf(\alpha_k, \alpha_j)$ to $Conf'$

Table 4.1: Conflict modification decision matrix.

Given the theory modification conditions to consider, we will now present the conflict modifying decision justification procedure. A more formal description of the procedure can be found in procedure A.3 (in the appendix). The conflict modifying decision justification procedure can be informally described as three basic steps. Given $AT = \langle AR, Conf, Pref \rangle$, we firstly check that S is conflict-free. Secondly, we check that every $\alpha_i \in AR$ that defeats some element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$. Thirdly, every $\alpha_i \in AR$ that is not in conflict with any element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$. Finally, we return the modified $Conf'$. In more detail, the conflict modifying decision justification procedure can be described as the following steps:

1. Firstly, we check and eliminate any conflict within S hence, for all $\alpha_i, \alpha_j \in S$, if $deft(\alpha_i, \alpha_j)$ then we remove the assertions $conf(\alpha_i, \alpha_j)$ and $conf(\alpha_j, \alpha_i)$ from $Conf'$.

2. To minimise the search space, we filter the set of arguments AR and S. To filter the set of arguments AR, we construct a set of arguments $AR' \subseteq AR$ by removing all arguments that are defeated by some $\alpha_i \in S$ from consideration and construct a set of arguments $S' \subseteq S$ consisting of arguments that defeat some $\alpha_j \in AR$. This is performed to eliminate situations shown in example 4.3.2.

3. Secondly, we check that every $\alpha_i \in AR'$ that defeats some element $\alpha_j \in S'$ is defeated by some element $\alpha_k \in S'$. To achieve this task, we perform the following:

(a) For each $deft(\alpha_i, \alpha_j) \in Deft$, if $\alpha_i \in AR'$ and $\alpha_k \in (S' \cup \{\alpha_j\})$ then non-deterministically select some element $\alpha_k \in S'$ and add the assertion $conf(\alpha_j, \alpha_i)$ to $Conf'$. Example 4.3.2 also shows that if we were to only consider undefeated arguments in S, it is possible that the set S' might be potentially empty. To resolve this issue, we will consider $S' \cup \{\alpha_j\}$ where $\alpha_j \in S$ is the argument currently involved in the defeat relationship.

(b) However, if we happened to select α_j, we need to check that α_j is not defeated by α_i based on the preference theory. If α_j is selected and it is not the case that $pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$ then we remove the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$.

(c) Since we have now defeated α_i, we can remove the argument α_i from AR'.

4. Lastly, we check that every $\alpha_i \in AR$ that is not in conflict with any element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$. Since arguments in $AR' \subseteq AR$ are the only arguments that have not been defeated by some $\alpha_j \in S$, we will only need to consider the arguments in AR'. To achieve this task, we perform the following:

(a) For each $\alpha_i \in AR'$, non-deterministically select some element $\alpha_k \in S$ and remove the assertion $conf(\alpha_i, \alpha_k)$ while adding the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. The removal of $conf(\alpha_i, \alpha_k)$ is a precautionary measure since there exists two instances where $deft(\alpha_i, \alpha_k) \notin Deft$ and $deft(\alpha_k, \alpha_i) not \in Deft$ occurs. Firstly, $conf(\alpha_i, \alpha_k) \notin Conf$ and $conf(\alpha_k, \alpha_i) \notin Conf$. However, if it is the case that $\{conf(\alpha_i, \alpha_k), conf(\alpha_k, \alpha_i)\} \subseteq Conf$, then it will be the case that $\{pref(\{\alpha_i\}, \{\alpha_k\}), pref(\{\alpha_k\}, \{\alpha_i\})\} \subseteq Pref$. Hence, in such an instance we are required to remove the assertion $conf(\alpha_i, \alpha_k)$ from $Conf'$.

Note that the conflict theory modification procedure can also be used as the alternative to the preferences theory modification in the situation where $S \in Ext_{AT}$ mentioned previously.

Example 4.3.2. *Given $AT = \langle AR, Conf, Pref \rangle$ and a user-specified decision S. Let us assume that $\alpha_i, \alpha_j, \alpha_k, \alpha_l \in AR$, $\alpha_i, \alpha_j \in S$, $\alpha_k, \alpha_l \notin S$ and $\{conf(\alpha_k, \alpha_i), conf(\alpha_l, \alpha_j)\} \subseteq Conf$. If we do not filter S and consider only arguments from S that are not defeated, the selection process will allow for the possibility of adding the assertion $conf(\alpha_i, \alpha_l)$ and $conf(\alpha_j, \alpha_k)$ into $Conf'$. In such a situation, we will not get S as the unique extension of AT'. Figure 4.6 presents a diagrammatic illustration of the issue.*
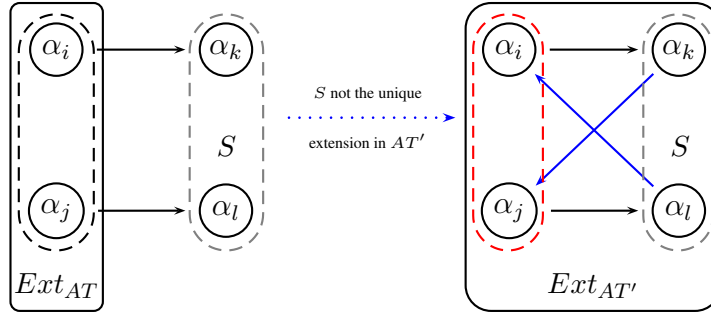
Figure 4.6: Graphical Illustration of Conflict Modification Exception

Properties of The Conflict modifying Decision Justification Procedure

We will now consider the termination, soundness and minimal change properties of the conflict modifying decision justification procedure. Firstly, let us consider the termination properties of the conflict modifying decision justification procedure. The termination property illustrates that the conflict modifying decision justification procedure will eventually halt.

Theorem 4.3.7 (Termination). *Given an argumentation theory* $AT = \langle AR, Conf, Pref \rangle$, *if* AR *and* S *are finite then the conflict modifying decision justification procedure terminates.*
Proof:
The conflict modifying decision justification procedure terminates if all loops terminate. Given that the conflict modifying decision justification procedure consists of three basic steps: Check that S is conflict-free, check that every $\alpha_i \in AR$ that defeats some element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$ and check that every $\alpha_i \in AR$ that is not in conflict with any element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$. This entails that conflict modifying decision justification procedure terminates if the three basic steps terminate.

Firstly, let us consider the first step. Given the procedure evaluates every $\alpha_i, \alpha_j \in S$ the procedure terminates when all $\alpha_i, \alpha_j \in S$ have been evaluated. Let us assume that S is finite and the procedure does not terminate. Given that S is a finite set, then eventually all $\alpha_i, \alpha_j \in S$ will be evaluated. Hence, the procedure terminates. Thus violating the assumption.

Secondly, let us consider the second step. Given the procedure evaluates every $\alpha_i \in AR$, the procedure terminates when all $\alpha_i \in AR$ have been evaluated. Let us assume that AR is finite and the procedure does not terminate. Given that AR is a finite set, then eventually all $\alpha_i \in AR$ will be evaluated. Hence, the procedure terminates. Thus violating the assumption.

Finally, let us consider the third step. Given the procedure evaluates every $\alpha_i \in AR$, the procedure terminates when all $\alpha_i \in AR$ have been evaluated. Let us assume that AR is finite

and the procedure does not terminate. Given that AR is a finite set, then eventually all $\alpha_i \in AR$ will be evaluated. Hence, the procedure terminates. Thus violating the assumption. $\quad\square$

Let us now consider the soundness property of the conflict modifying decision justification procedure. The soundness property illustrates that the procedure will always return a modified conflict theory such that the user-specified decision is the unique decision.

Theorem 4.3.8 (Soundness). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$ a (user-specified) decision S and S is not the unique decision of AT, the conflict modifying decision justification procedure is sound.*

Proof:
The conflict modifying decision justification procedure is sound if and only if the conflict modifying decision justification procedure realises a conflict modifying decision justification function. In other words, the conflict modifying decision justification procedure returns a $Conf'$ such that S is the unique decision of $AT' = \langle AR, Conf', Pref \rangle$. S is the unique decision of $AT' = \langle AR, Conf', Pref \rangle$ if and only if it satisfies the following conditions:

1. There does not exist $\alpha_i, \alpha_j \in S$ such that $deft(\alpha_i, \alpha_j) \in Deft$. Let us assume that there exists $\alpha_i, \alpha_j \in S$ such that $deft(\alpha_i, \alpha_j) \in Deft$ and S is the unique decision. This entails that there exists $\alpha_i, \alpha_j \in S$ such that $conf(\alpha_i, \alpha_j) \in Conf'$. However, the first step of the conflict modifying decision justification procedure checks that S is conflict-free. Hence, for all $\alpha_i, \alpha_j \in S$, if $conf(\alpha_i, \alpha_j) \in Conf'$ then, the assertion $conf(\alpha_i, \alpha_j)$ is removed from $Conf'$. Therefore it is not the case that $conf(\alpha_i, \alpha_j) \in Conf'$, thus violating the assumption.

2. For all $\alpha_i \in AR$, if $\alpha_i \notin S$ then there exists $\alpha_j \in S$ such that $deft(\alpha_j, \alpha_i) \in Deft$. Let us assume that $\alpha_i \notin S$ and there does not exist $\alpha_j \in S$ such that $deft(\alpha_j, \alpha_i) \in Deft$. Let us consider two cases:

Case 1: Let us assume that $conf(\alpha_j, \alpha_i) \notin Conf'$. However, for every $\alpha_i \in AR$ that is not defeated by some element $\alpha_k \in S$, the second and third steps of the conflict modifying decision justification procedure add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. Hence, it must be the case that there exists some element $\alpha_j \in S$ such that $conf(\alpha_j, \alpha_i) \in Conf'$. Thus violating the assumption.

Case 2: Let us assume that $\{conf(\alpha_j, \alpha_i), conf(\alpha_i, \alpha_j)\} \subseteq Conf'$ and $pref(\{\alpha_i\}, \{\alpha_j\}) \in Pref$. However, for every $\alpha_i \in AR$ that is not defeated by some element $\alpha_j \in S$, if it is the not the case that $pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref'$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$, the second and third steps of the conflict modifying decision justification

procedure removes the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$. Hence, it cannot be the case that there exists $conf(\alpha_i, \alpha_j) \in Conf'$. Thus violating the assumption. $\square$

Let us now consider the existence property of the conflict modifying decision justification function. The existence property illustrates that there exists at least one conflict modifying decision justification function that will be able to perform the required modification such that the user-specified decision is the unique decision of the modified argumentation theory.

Theorem 4.3.9 (Existence). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S, if S is not the unique decision of AT, then there exists at least one conflict modifying decision justification function such that S is the unique decision of $A_{just}(\langle AR, Conf, Pref \rangle, S)$*

Proof:
Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S, if S is not the unique decision of AT, then there exists $AT' = \langle AR, Conf', Pref \rangle$ such that S is the unique decision. Since there exists $AT' = \langle AR, Conf', Pref \rangle$ such that S is the unique decision then there exists a conflict modifying decision justification function such that S is the unique decision of $A_{just}(\langle AR, Conf, Pref \rangle, S)$. $\square$

We will now consider the minimally modification property of the conflict modifying decision justification function. Given a modified conflict theory $Conf'$ derived from $Conf$. We consider $Conf'$ to be a minimally modified conflict theory if there does not exist any subset of $Conf'$ that will yield the same unique decision user-specified decision.

Definition 4.3.2. *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$ and a (user-specified) decision $S \subseteq AR$, a conflict theory $Conf'$ is a minimally modified conflict theory with respect to AT and S if and only if:*

1. *S is the unique decision of $\langle AR, Conf', Pref \rangle$.*

2. *There exists no $Conf''$ such that $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and S is the unique decision of $\langle AR, Conf'', Pref \rangle$.*

Note that $\ominus$ denotes symmetric set difference.

Note that a minimal change to a conflict theory can also be expressed with respect to set cardinality by using $|Conf \ominus Conf''| < |Conf \ominus Conf'|$. From the definition above, we

can assume that starting with $conf(\alpha_i, \alpha_j)$, both removing $conf(\alpha_i, \alpha_j)$ from $Conf'$ and replacing $conf(\alpha_i, \alpha_j)$ with $conf(\alpha_j, \alpha_i)$ represent the same "quantum of change".

Let us now consider the minimality property of the conflict-modifying decision justification function. The minimality property illustrates that the conflict modifying decision justification procedure will produce a result that is a minimal repair to the original conflict theory such that the user-specified decision is the unique decision.

Theorem 4.3.10 (Minimal). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S and S is not the unique decision of AT, the conflict modifying decision justification procedure generates a conflict theory $Conf'$ that is a minimally modified conflict theory with respect to AT and S. As well $\langle AR, Conf', Pref \rangle$ is a minimally (w.r.t set inclusion and set cardinality) modified argumentation theory with respect to AT and S.*

Proof:

$Conf'$ is a minimally (w.r.t set inclusion and set cardinality) modified conflict theory with respect to AT and S if and only if it satisfies the following properties:

1. S is the unique decision of $\langle AR, Conf', Pref \rangle$. Let us assume that S is not the unique decision of $\langle AR, Conf', Pref \rangle$. This entails that the conflict modifying decision justification procedure is unsound and returns modified conflict theory $Conf'$ such that S is not the unique decision. However, theorem 4.3.8 shows that the conflict modifying decision justification procedure is sound and will always return $Conf'$ such that S is the unique decision of $\langle AR, Conf', Pref \rangle$. Thus violating the assumption.

2. There exists no $Conf''$ such that $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and S is the unique decision of $AT'' = \langle AR, Conf'', Pref \rangle$. Let us assume that there exists a $Conf''$ such that $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and S is the unique decision of AT''. This entails that there exists $\alpha_i \in AR$ and there does not exist any $\alpha_j \in S$ such that $conf(\alpha_j, \alpha_i) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf''$. It then follows that there exists $\alpha_i \in AR$ that is not defeated by any $\alpha_j \in S$. Hence, S is not an extension of AT''. Thus violating the assumption.

3. There exists no $Conf''$ such that $|Conf \ominus Conf''| < |Conf \ominus Conf'|$ and S is the unique decision of $AT'' = \langle AR, Conf'', Pref \rangle$. Let us assume that there exists a $Conf''$ such that $|Conf \ominus Conf''| < |Conf \ominus Conf'|$ and S is the unique decision of AT''. This entails that there exists $\alpha_i \in AR$ and there does not exist any $\alpha_j \in S$ such that $conf(\alpha_j, \alpha_i) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf''$. It then follows that there exists $\alpha_i \in AR$ that is not defeated by any $\alpha_j \in S$. Hence, S is not an extension of AT''. Thus violating the assumption. $\square$

4.3.3 Preference & Conflict Modifying Decision Justification Procedure

The preference modifying decision justification procedure and the conflict modifying decision justification procedure only modify the respective theories. The **combined conflict modifying and preference modifying decision justification procedure** interleaves the modification of preference and conflict theories. The underlying principle deployed in the combined conflict modifying and preference modifying decision justification procedure is similar to that of the conflict modifying decision justification procedure. However, the combined conflict modifying and preference modifying decision justification procedure allows users the choice of modifying either the preference theory, conflict theory or an interleaving of both. Hence, providing the user with a greater degree of freedom. Given $AT = \langle AR, Conf, Pref \rangle$, Ext_{AT} and a user-specified decision S, the combined modification of the conflict theory and preference theory aims at promoting S to being the only extension in Ext_{AT}. To achieve this, all arguments $\alpha_i \in AR \setminus S$ need to be defeated by some $\alpha_j \in S$. The conflict and preference relationship between each $\alpha_i \in AR \setminus S$ and S maybe be characterised as one of several possibilities illustrated in Figure 4.5. Firstly, we determine if $\{conf(\alpha_i, \alpha_j), conf(\alpha_j, \alpha_i) \subseteq Conf$. If it is the case that $conf(\alpha_i, \alpha_j) \notin Conf$ and $conf(\alpha_j, \alpha_i) \in Conf$ then there is no modification required. Figure 4.5a illustrates such a situation. If it is the case that $conf(\alpha_i, \alpha_j), conf(\alpha_j, \alpha_i) \subseteq Conf$, then we determine if $\{pref(\{\alpha_i\}, \{\alpha_j\}), pref(\{\alpha_j\}, \{\alpha_i\})\} \subseteq Pref$. If $pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$, then there is no modification required. Figure 4.5b illustrates such a situation. However, in the instance where $\{pref(\{\alpha_i\}, \{\alpha_j\}), pref(\{\alpha_j\}, \{\alpha_i\})\} \subseteq Pref$, one of three possible modifications can be exclusively performed. We can either:

- Remove the assertion $pref(\{\alpha_i\}, \{\alpha_j\})$ from $Pref'$ or

- Remove the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$ or

- Select some $\alpha_k \in S$ that is not defeated and add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$

Figure 4.5d, 4.5e and 4.5f illustrate such a situation. In the instance that $pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \notin Pref$, one of three possible modifications can be exclusively performed. We can either:

- Remove the assertion $pref(\{\alpha_i\}, \{\alpha_j\})$ from $Pref'$ and add the assertion $pref(\{\alpha_j\}, \{\alpha_i\})$ to $Pref'$ or

- Remove the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$ or

- Select some $\alpha_k \in S$ that is not defeated and add the assertion α_k, α_i to $Conf'$

Figure 4.5e, 4.5f and 4.5g illustrate such a situation. If it is the case that $conf(\alpha_i, \alpha_j) \in Conf$ and $conf(\alpha_j, \alpha_i) \notin Conf$, then we need to select some $k \in S$ and add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. Furthermore, if $pref(\{\alpha_i\}, \{\alpha_k\}) \in Pref$, we need to remove the assertion $conf(\alpha_i, \alpha_k)$ from $Conf'$. Figure 4.5h, 4.5i and 4.5j illustrate such a situation. Finally, if it is the case that $conf(\alpha_i, \alpha_j) \notin Conf$ and $conf(\alpha_j, \alpha_i) \notin Conf$, then we need to select some $k \in S$ and add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. Figure 4.5c illustrates such a situation. Table 4.2 provides a summary of the conflict theory modification decision matrix.

List of cases to consider for any $x, y \in AR$ and $y \in S$				
$Conf$		$Pref$		
$conf(\alpha_i, \alpha_j)$	$conf(\alpha_j, \alpha_i)$	$pref(\{\alpha_i\}, \{\alpha_j\})$	$pref(\{\alpha_j\}, \{\alpha_i\})$	Solutions
True	False	True	False	(2 & 3) or 6 or (2 & 4 & 5)
True	False	False	True	2 or 6
True	False	False	False	(2 & 3) or (2 & 4) or 6
True	False	True	True	(2 & 3) or (2 & 5) or 6
True	True	True	False	3 or 6 or (4 & 5)
True	True	False	True	1
True	True	False	False	(2 & 3) or 4 or 6
True	True	True	True	(2 & 3) or 5 or 6
False	False	-	-	6
False	True	-	-	1

List of solutions
1. Do Nothing
2. Add the assertion $conf(\alpha_j, \alpha_i)$ into $Conf'$
3. Remove the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$
4. Add $pref(\{\alpha_j\}, \{\alpha_i\})$ into $Pref'$
5. Remove the assertion $pref(\{\alpha_i\}, \{\alpha_j\})$ from $Pref'$
6. Select some $\alpha_k \in S$ that is not defeated and add the assertion $conf(\alpha_k, \alpha_j)$ to $Conf'$

Table 4.2: Preference & Conflict modification decision matrix.

Given the theory modification conditions to consider, we will now present the combined conflict modifying and preference modifying decision justification procedure. A more formal description of the procedure can be found in procedure A.4 (in the appendix). Similar to the conflict modifying decision justification procedure, the combined conflict modifying and preference modifying decision justification procedure can be informally described as three

basic steps. Firstly, we ensure that S is conflict-free. Secondly, we ensure that every $\alpha_i \in AR$ that defeats some element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$ by modifying either the conflict theory, preference theory or an interleaving of both. Thirdly, we ensure that every $\alpha_i \in AR$ that is not in conflict with any element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$ by modifying either the conflict theory, preference theory or both. Finally, we return the modified $Pref'$ and $Conf'$. In more detail, the combined conflict modifying and preference modifying decision justification procedure can be described as the following steps:

1. Firstly, we check and eliminate any conflict within S hence, for all $\alpha_i, \alpha_j \in S$, if $deft(\alpha_i, \alpha_j) \in Deft$, then we check if $\{conf(\alpha_i, \alpha_j), conf(\alpha_j, \alpha_i)\} \subseteq Conf$. If $\{conf(\alpha_i, \alpha_j), conf(\alpha_j, \alpha_i)\} \not\subseteq Conf$ then, we remove the assertion $conf(\alpha_i, \alpha_j)$ and $conf(\alpha_j, \alpha_i)$ from $Conf'$ else we perform *one* of the following:

   (a) We remove the assertion $conf(\alpha_i, \alpha_j)$ and $conf(\alpha_j, \alpha_i)$ from $Conf'$.

   (b) We add the assertions $pref(\{\alpha_i\}, \{\alpha_j\})$ and $pref(\{\alpha_j\}, \{\alpha_i\})$ to $Pref'$.

2. To minimise the search space, we filter the set of arguments AR and S. To filter the set of arguments AR, we construct a set of argument $AR' \subseteq AR$ by removing from consideration, all arguments that are defeated by some $\alpha_i \in S$ and constructing a set of arguments $S' \subseteq S$ consisting of arguments that defeated some $\alpha_j \in AR$. This is performed to eliminate situations shown in example 4.3.2.

3. Secondly, we check that every $\alpha_i \in AR'$ that defeats some element $\alpha_j \in S'$ is defeated by some element $\alpha_k \in S'$. To achieve this task, we perform the following:

   (a) For each $deft(\alpha_i, \alpha_j) \in Deft$, if $\alpha_i \in AR'$ and $\alpha_k \in (S' \cup \{\alpha_j\})$ then select some element $\alpha_k \in S'$ and add the assertion $conf(\alpha_j, \alpha_i)$ to $Conf'$. Example 4.3.2 also shows that if we were to only consider undefeated arguments in S, it is possible that the set S' might be potentially empty. To resolve this issue, we will consider $S' \cup \{\alpha_j\}$ where $\alpha_j \in S$ is the argument currently involved in the defeat relationship.

   (b) However, if we happened to select α_j, we need to check that α_j is not defeated by α_i based on the preference theory. If α_j is selected and it is not the case that $pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$ then we perform *one* of the following:

      i. We remove the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$.

      ii. We remove the assertion $pref(\{\alpha_i\}, \{\alpha_j\})$ and add the assertion $pref(\{\alpha_j\}, \{\alpha_i\})$ to $Pref'$.

(c) Since we have now defeated α_i, we can remove the argument α_i from AR'.

4. Lastly, we check that every $\alpha_i \in AR$ that is not in conflict with any element $\alpha_j \in S$ is defeated by some element $\alpha_k \in S$. Since arguments in $AR' \subseteq AR$ are the only arguments that have not been defeated by some $\alpha_j \in S$, we will only need to consider the arguments in AR'. To achieve this task, we perform the following:

(a) For each $\alpha_i \in AR'$, we select some element $\alpha_k \in S$. if $\{conf(\alpha_i, \alpha_k), conf(\alpha_k, \alpha_i)\} \not\subseteq Conf'$ then, we remove the assertion $conf(\alpha_i, \alpha_k)$ and add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$ else we perform *one* of the following:

   i. We remove the assertion $conf(\alpha_i, \alpha_k)$ from $Conf'$.
   ii. We add the assertions $pref(\{\alpha_i\}, \{\alpha_k\})$ to $Pref'$.

Properties of The Combined Preference & Conflict Modifying Decision Justification Procedure

We will now consider the termination, soundness and minimal change properties of the preference & conflict modifying decision justification procedure. Firstly, let us consider the termination properties of the preference & conflict modifying decision justification procedure. The termination property illustrates that the preference & conflict modifying decision justification procedure will eventually halt.

Theorem 4.3.11 (Termination). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, if AR and S are finite then the combined preference and conflict modifying decision justification procedure terminates.*

Proof:
The combined preference and conflict modifying decision justification procedure terminates if all loops terminate. Given that the combined preference and conflict modifying decision justification procedure consists of three basic steps: This entails that conflict modifying decision justification procedure terminates if the three basic steps terminate.

Firstly, let us consider the first step. Given the procedure evaluates every $\alpha_i, \alpha_j \in S$ the procedure terminates when all $\alpha_i, \alpha_j \in S$ have been evaluated. Let us assume that S is finite and the procedure does not terminate. Given that S is a finite set, then eventually all $\alpha_i, \alpha_j \in S$ will be evaluated. Hence, the procedure terminates. Thus violating the assumption.

Secondly, let us consider the second step. Given the procedure evaluates every $\alpha_i \in AR$, the procedure terminates when all $\alpha_i \in AR$ have been evaluated. Let us assume that AR is

finite and the procedure does not terminate. Given that AR is a finite set, then eventually all $\alpha_i \in AR$ will be evaluated. Hence, the procedure terminates. Thus violating the assumption.

Finally, let us consider the third step. Given the procedure evaluates every $\alpha_i \in AR$, the procedure terminates when all $\alpha_i \in AR$ have been evaluated. Let us assume that AR is finite and the procedure does not terminate. Given that AR is a finite set, then eventually all $\alpha_i \in AR$ will be evaluated. Hence, the procedure terminates. Thus violating the assumption.   □

Let us now consider the soundness property of the preference & conflict modifying decision justification procedure. The soundness property illustrates that the procedure will always return a modified conflict theory such that the user-specified decision is the unique decision.

Theorem 4.3.12 (Soundness). *Given $AT = \langle AR, Conf, Pref \rangle$ and a decision $S \subseteq AR$, if S is not a decision of AT, then preference and conflict modifying decision justification procedure returns $Conf'$ and $Pref'$ such that S is a decision of $\langle AR, Conf', Pref' \rangle$*

Proof:
The preference & conflict modifying decision justification procedure is sound if and only if the preference & conflict modifying decision justification procedure realises a preference & conflict modifying decision justification function. In other words, the preference & conflict modifying decision justification procedure a $Pref'$ and $Conf'$ such that S is the unique decision of $AT' = \langle AR, Conf', Pref' \rangle$. S is the unique decision of $AT' = \langle AR, Conf', Pref' \rangle$ if and only if it satisfies the following conditions:

1. There does not exist $\alpha_i, \alpha_j \in S$ such that $deft(\alpha_i, \alpha_j) \in Deft$. Let us assume that there exists $\alpha_i, \alpha_j \in S$ such that $deft(\alpha_i, \alpha_j) \in Deft$ and S is the unique decision. This entails that there exists one of the following possibilities:

Case 1: Let us assume that there exists $\alpha_i, \alpha_j \in S$ such that $conf(\alpha_i, \alpha_j) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf'$. However, the first step of the preference & conflict modifying decision justification procedure checks that S is conflict-free. Hence, for all $\alpha_i, \alpha_j \in S$, if $conf(\alpha_i, \alpha_j) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf'$ then, the assertion $conf(\alpha_i, \alpha_j)$ is removed from $Conf'$. This entails that it is not the case that $deft(\alpha_i, \alpha_j) \in Deft$, thus violating the assumption.

Case 2: Let us assume that there exists $\alpha_i, \alpha_j \in S$ such that $\{conf(\alpha_i, \alpha_j), conf(\alpha_j, \alpha_i)\} \subseteq Conf'$ and $\{pref(\{\alpha_i\}, \{\alpha_j\}), pref(\{\alpha_j\}, \{\alpha_i\})\} \not\subseteq Pref'$. However, the first step of the preference & conflict modifying decision justification procedure checks that S is conflict-free. Hence, for all $\alpha_i, \alpha_j \in S$, if $\{conf(\alpha_i, \alpha_j), conf(\alpha_j, \alpha_i)\} \subseteq$

$Conf'$ and $\{pref(\{\alpha_i\}, \{\alpha_j\}), pref(\{\alpha_j\}, \{\alpha_i\})\} \not\subseteq Pref'$ then, either both assertions are removed from $Conf'$ or the assertions $pref(\{\alpha_i\}, \{\alpha_j\})$ and $pref(\{\alpha_j\}, \{\alpha_i\})$ are added to $Pref'$. This entails that it is not the case that $deft(\alpha_i, \alpha_j) \in Deft$, thus violating the assumption.

2. For all $\alpha_i \in AR$, if $\alpha_i \notin S$ then there exists $\alpha_j \in S$ such that $deft(\alpha_j, \alpha_i) \in Deft$. Let us assume that $\alpha_i \notin S$ and there does not exist $\alpha_j \in S$ such that $deft(\alpha_j, \alpha_i) \in Deft$. Let us consider two cases:

Case 1: Let us assume that $conf(\alpha_j, \alpha_i) \notin Conf'$. However, for every $\alpha_i \in AR$ that is not defeated by some element $\alpha_k \in S$, the second and third steps of the preference & conflict modifying decision justification procedure add the assertion $conf(\alpha_k, \alpha_i)$ to $Conf'$. Hence, it must be the case that there exists some element $\alpha_j \in S$ such that $conf(\alpha_j, \alpha_i) \in Conf'$. Thus violating the assumption.

Case 2: Let us assume that $\{conf(\alpha_j, \alpha_i), conf(\alpha_i, \alpha_j)\} \subseteq Conf'$ and $pref(\{\alpha_i\}, \{\alpha_j\}) \in Pref$. However, for every $\alpha_i \in AR$ that is not defeated by some element $\alpha_j \in S$, if it is the not the case that $pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref'$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref$, the second and third steps of the preference & conflict modifying decision justification procedure removes the assertion $conf(\alpha_i, \alpha_j)$ from $Conf'$ or removes the assertion. $pref(\{\alpha_i\}, \{\alpha_j\})$ and add the assertion $pref(\{\alpha_j\}, \{\alpha_i\})$ into $Pref'$. Hence, it cannot be the case that there exists $conf(\alpha_i, \alpha_j) \in Conf'$ or $pref(\{\alpha_i\}, \{\alpha_j\}) \in Pref'$. Thus violating the assumption. $\square$

Let us now consider the existence property of the preference & conflict modifying decision justification function. The existence property illustrates that there exists at least one preference & conflict modifying decision justification function that will be able to perform the required modification such that the user-specified decision is the unique decision of the modified argumentation theory.

Theorem 4.3.13 (Existence). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S, if S is not the unique decision of AT, then there exists at least one preference & conflict modifying justification function such that S is the unique decision of $A_{just}(\langle AR, Conf, Pref \rangle, S)$*

Proof:

Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a (user-specified) decision S, if S is not the unique decision of AT, then there exists $AT' = \langle AR, Conf', Pref' \rangle$ such that S is the unique decision. Since there exists $AT' = \langle AR, Conf', Pref' \rangle$ such that S is the unique decision then there exists a conflict modifying decision justification function such that S is the unique decision of $A_{just}(\langle AR, Conf, Pref \rangle, S)$. $\square$

We will now consider the minimally modification property of the preference & conflict modifying decision justification procedure function. Given a modified preference theory $Pref'$ and modified conflict theory $Conf'$ derived from $Pref$ and $Conf$ respectively, we consider $Pref'$ and $Conf'$ to be a minimally modified preference & conflict theory if there does not exist any subset of $Pref'$ and $Conf'$ that will yield the same unique user-specified decision.

Definition 4.3.3. *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, and a (user-specified) decision $S \subseteq AR$, $\langle AR, Conf', Pref' \rangle$ is a minimally modified argumentation theory if and only if:*

1. *S is the unique decision of $\langle AR, Conf', Pref' \rangle$*

2. *There exists no $Conf''$ and $Pref''$ such that $Conf \ominus Conf'' \subset Conf \ominus Conf'$, $Pref \ominus Pref'' \subset Pref \ominus Pref'$ and S is the unique decision $\langle AR, Conf'', Pref'' \rangle$*

Note that $\ominus$ denotes symmetric set difference.

Note that a minimal change to a preference & conflict theory can also be expressed with respect to set cardinality by using $|Pref \ominus Pref''| < |Pref \ominus Pref'|$ and $|Conf \ominus Conf''| < |Conf \ominus Conf'|$. From the definition above, we can assume that starting with $conf(\alpha_i, \alpha_j)$, both removing $conf(\alpha_i, \alpha_j)$ from $Conf'$ and replacing $conf(\alpha_i, \alpha_j)$ with $conf(\alpha_j, \alpha_i)$ represent the same "quantum of change".

Let us now consider the minimality property of the preference & conflict modifying decision justification function. The minimality property illustrates that the preference & conflict modifying decision justification procedure will produce a result that is a minimal repair to the original conflict theory such that the user-specified decision is the unique decision.

Theorem 4.3.14 (Minimal). *Given an argumentation theory $AT = \langle AR, Conf, Pref \rangle$ and a (user-specified) decision and S is not the unique decision of AT, the preference & conflict modifying decision justification procedure generates a preference theory $Pref'$ and conflict theory $Conf'$ that is a minimally modified preference and conflict theory with respect to AT and S. As well $\langle AR, Conf', Pref' \rangle$ is a minimally (w.r.t set inclusion and set cardinality) modified argumentation theory with respect to AT and S.*

Proof:
$Pref'$ and $Conf'$ are minimally (w.r.t set inclusion and set cardinality) modified preference & conflict theory with respect to AT and S if and only if it satisfies the following properties:

1. S is the unique decision of $\langle AR, Conf', Pref' \rangle$. Let us assume that S is not the unique decision of $\langle AR, Conf', Pref' \rangle$. This entails that the preference & conflict modifying decision justification procedure is unsound and returns modified preference theory $Pref'$ and conflict theory $Conf'$ such that S is not the unique decision. However, theorem 4.3.12 shows that the preference & conflict modifying decision justification procedure is sound and will always return $Pref'$ and $Conf'$ such that S is the unique decision of $\langle AR, Conf', Pref' \rangle$. Thus violating the assumption.

2. There exists no $Pref''$ and $Conf''$ such that $Pref \ominus Pref'' \subset Pref \ominus Pref'$ and $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and S is the unique decision of $AT'' = \langle AR, Conf'', Pref'' \rangle$. Let us assume that there exists $Pref''$ and $Conf''$ such that $Pref \ominus Pref'' \subset Pref \ominus Pref'$, $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and S is the unique decision of AT''. This entails that there exists $\alpha_i \in AR$ and there does not exist any $\alpha_j \in S$ such that $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref'$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \notin Pref''$ and $conf(\alpha_j, \alpha_i) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf''$. It then follows that there exists $\alpha_i \in AR$ that is not defeated by any $\alpha_j \in S$. Hence, S is not an extension of AT''. Thus violating the assumption.

3. There exists no $Pref''$ and $Conf''$ such that $|Pref \ominus Pref''| < |Pref \ominus Pref'|$ and $|Conf \ominus Conf''| < |Conf \ominus Conf'|$ and S is the unique decision of $AT'' = \langle AR, Conf'', Pref'' \rangle$. Let us assume that there exists a $Pref''$ and $Conf''$ such that $|Pref \ominus Pref''| < |Pref \ominus Pref'|$, $|Conf \ominus Conf''| < |Conf \ominus Conf'|$ and S is the unique decision of AT''. This entails that there exists $\alpha_i \in AR$ and there does not exist any $\alpha_j \in S$ such that $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref'$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \notin Pref''$ and $conf(\alpha_j, \alpha_i) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf''$. It then follows that there exists $\alpha_i \in AR$ that is not defeated by any $\alpha_j \in S$. Hence, S is not an extension of AT''. Thus violating the assumption. $\qquad \square$

4.3.4 General Decision Justification Procedure

We will now describe the **general decision justification procedure** informally. A more formal description of the procedure can be found in procedure A.1 (in the appendix). During the decision justification interaction, any combination of three categories of change may occur: the addition of new arguments, the modification of the conflict theory or the modification of the preference theory. Given $AT = \langle AR, Conf, Pref \rangle$ and a user-specified decision S, if S is the unique decision then the decision justification phase terminates. Otherwise, we check if AR is S-complete. If AR is not S-complete, we preform $AR \cup S$ and execute decision

generation procedure. In [35, 36], the authors addressed the issue of revising the set of arguments focusing on minimal extension revision, hence in situations where we are interested in minimal extension revision or no preference theory revision is required, the techniques they described can be deployed.

If AR is S-complete, we determine if $S \in Ext_{AT}$. This provides useful information to assist in determining which of the two categories (conflict theory or preferences theory) of modification to perform next. If $S \in Ext_{AT}$, this informs us that a change in the preference theory such that there is no decision $S' \in Ext_{AT}$ and S' is more preferred to S. If $S \notin Ext_{AT}$, we modify the conflict theory such that S is the unique decision of $AT' = \langle AR, Conf', Pref \rangle$. Hence, if $S \in Ext_{AT}$, then we execute the **preference modifying decision justification procedure**. If $S \notin Ext_{AT}$, then we execute the **conflict modifying decision justification procedure**. After each modification, the generation procedure will need to be executed to check the validity of the justifications.

Properties of The General Decision Justification Procedure

We will now consider the termination, soundness and minimal change properties of the general decision justification procedure. Firstly, let us consider the termination properties of the general decision justification procedure. The termination property illustrates that the general decision justification procedure will eventually halt.

Theorem 4.3.15 (Termination). *The general decision justification procedure terminates.*

Proof:
The general decision justification procedure terminates if and only if it satisfies the following conditions:

- Preference modifying decision justification procedure terminates.

- Preference modifying decision justification procedure is sound.

- Conflict modifying decision justification procedure terminates.

- Conflict modifying decision justification procedure is sound.

From theorem 4.3.1, it follows that the preference modifying decision justification procedure terminates. From theorem 4.3.2, it follows that the preference modifying decision justification procedure is sound. From theorem 4.3.7, it follows that the conflict modifying decision

justification procedure terminates. From theorem 4.3.8, it follows that the conflict modifying decision justification procedure is sound. Hence general decision justification procedure terminates. □

Let us now consider the soundness property of the general decision justification procedure. The soundness property illustrates that the procedure will always return a modification to the original argumentation theory such that the user-specified decision is the unique decision.

Theorem 4.3.16 (Soundness). *Given* $AT = \langle AR, Conf, Pref \rangle$, *a (user-specified) decision* S *and* S *is not the unique decision of* AT, *the general decision justification procedure is sound.*

Proof:
Let us assume that the general decision justification procedure is not sound. This entails that either the preference modifying decision justification procedure or conflict modifying decision justification procedure is not sound. Let us consider two cases:

Case 1: Preference modifying decision justification procedure is not sound. This entails that preference modifying decision justification procedure does not realise a preference modifying decision justification function. However, theorem 4.3.2 shows that preference modifying decision justification procedure does realise a preference modifying decision justification function and therefore is sound. Thus, the assumption is violated.

Case 2: Conflict modifying decision justification procedure is not sound. This entails that conflict modifying decision justification procedure does not realise a conflict modifying decision justification function. However, theorem 4.3.8 shows that conflict modifying decision justification procedure does realise a conflict modifying decision justification function and therefore is sound. Thus, the assumption is violated. □

In the following section, we will firstly propose some properties for the theory change operators for the mixed-initiative argumentation framework. Secondly, we will perform a discussion on the flexibility of a mixed-initiative argumentation framework.

4.4 Discussion

In the previous section, we introduced the mixed-initiative argumentation framework ($MIAF$) and illustrated the ability for the framework to revise the underlying argumentation theory. Our aim in this section is to firstly to propose some fundamental properties governing the argumentation theory change operation. Secondly, we will present a discussion on the capabilities of such a framework focusing on the different combination of different structure in arguments and the presence/absence of a guiding knowledge-base.

4.4.1 Properties For Argumentation Theory Change

The AGM framework [3] provides an important basis to understanding theory change. Inspired by the AGM properties, we can discuss certain properties that should govern the kind of argumentation theory change that we are interested in (note that our operators does not easily fall into the traditional categories of revision, contraction or expansion). Assume an argumentation theory $AT = \langle AR, Conf, Pref \rangle$, a decision $S \subseteq AR$, a decision generation function A_{gen} and a decision justification function A_{just}. For each of the theory change operators, the first property articulates the principle of categorical matching which requires, in our instance, that the result of an argumentation theory change should be an argumentation theory. The second property behaves like the AGM *success* postulate. The third properties ensures that the conflict and preference theories respectively are minimally modified. The fourth property behaves in a similar fashion to the AGM *vacuity* postulate. The fifth property behaves in a similar fashion to the AGM *recovery* postulate.

Properties For Preference Theory Change

Let us consider the set of rationality properties governing the preference modifying decision justification function. Given $AT = \langle AR, Conf, Pref \rangle$, a decision $S \subseteq AR$, a decision generation function A_{gen} and a decision justification function A_{just}. The condition for the use of the preference modifying decision justification procedure is that $S \in Ext_{AT}$. Given this condition, the rationality properties are:

P1 $A_{just}(\langle AR, Conf, Pref \rangle, S)$ is an argumentation theory.

P2 $S = A_{gen}(A_{just}(AT, S))$

P3 $A_{just}(\langle AR, Conf, Pref \rangle, S) = \langle AR, Conf, Pref' \rangle$ s.t. there exists no $Pref''$ where $Pref \ominus Pref'' \subset Pref \ominus Pref'$

P4 If $S = A_{gen}(AT)$ then $A_{just}(AT, S) = AT$

P5 $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S)) = A_{gen}(A_{just}(AT, S))$

Property P1 articulates the principle of categorical matching which requires that the result of an argumentation theory change should be an argumentation theory. Property P2 behaves like the AGM *success* postulate. Property P3 ensures that the preference theories is minimally modified. Property P4 behaves in a similar fashion to the AGM *vacuity* postulate. Property P5 behaves in a similar fashion to the AGM *recovery* postulate.

Theorem 4.4.1. *A mixed-initiative argumentation system $\langle A_{gen}, A_{just} \rangle$, where A_{just} is an implementation of the preference modifying decision justification procedure described in section 4.3.1 satisfies (P1) - (P5).*

Proof:
Given that A_{just} is an implementation of the preference modifying decision justification procedure described in section 4.3.1, let us consider the following properties:

P1: $A_{just}(\langle AR, Conf, Pref \rangle, S)$ is an argumentation theory. The principle of categorical matching: Let us assume that A_{just} does not satisfy P1. This entails that A_{just} does not return an argumentation theory. Hence A_{just} is not a preference modifying decision justification function. Thus, the assumption is violated.

P2: $S = A_{gen}(A_{just}(AT, S))$: Let us assume that A_{just} does not satisfy P2. This entails that A_{just} does not return $AT' = \langle AR, Conf, Pref' \rangle$ such that S is the unique decision. Hence by definition (Definition 4.2.6), A_{just} is not preference modifying decision justification function. Thus, the assumption is violated.

P3: $A_{just}(\langle AR, Conf, Pref \rangle, S) = \langle AR, Conf, Pref' \rangle$ s.t. there exists no $Pref''$ where $Pref \ominus Pref'' \subset Pref \ominus Pref'$: Let us assume that A_{just} does not satisfy P3. This entails that there exists $Pref''$ where $Pref \ominus Pref'' \subset Pref \ominus Pref'$. This entails that there exists $A_i \in (Ext_{AT} \setminus \{S\})$ s.t. $pref(S, A_i) \in Pref'$ and $pref(S, A_i) \notin Pref''$. It then follows that there exists $A_i \in Ext_{AT''}$ that is as preferred as S. Hence, S is not a unique extension of AT''. Hence by definition (Definition 4.2.6), A_{just} is not preference modifying decision justification function. Thus, the assumption is violated.

P4: If $S = A_{gen}(AT)$ then $A_{just}(AT, S) = AT$: Let us assume that A_{just} does not satisfy P4. This entails that A_{just} does not return AT. It then follows that A_{just} returns a

modified $AT' = \langle AR, Conf, Pref' \rangle$ where $Pref' \neq Pref$. However theorem 4.3.4 shows that A_{just} will perform minimal change to the AT. Since $S = A_{gen}(AT)$ then there is no change to AT. Hence, it must be the case that $Pref' = Pref$. Thus, the assumption is violated.

P5: $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S)) = A_{gen}(A_{just}(AT, S))$: Let us assume that A_{just} does not satisfy P5. This entails that $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S))$ does not return S. It then follows that A_{just} is not sound. However theorem 4.3.2 shows that A_{just} is sound. Thus, the assumption is violated. $\qquad \square$

Properties For Conflict Theory Change

The conflict modifying decision justification procedure described in the previous section (section 4.3.2) utilises a non-deterministic selection function. Given a space of possible choices, we can describe a class of selection operators where each operator entails a unique conflict modifying decision justification procedure. Hence, the conflict modifying decision justification function describes a class of selection operators.Let us consider the set of rationality properties governing the preference modifying decision justification function. Given $AT = \langle AR, Conf, Pref \rangle$, a decision $S \subseteq AR$, a decision generation function A_{gen} and a decision justification function A_{just}. The rationality properties are:

C1 $A_{just}(\langle AR, Conf, Pref \rangle, S)$ is an argumentation theory.

C2 $S = A_{gen}(A_{just}(AT, S))$

C3 $A_{just}(\langle AR, Conf, Pref \rangle, S) = \langle AR, Conf', Pref \rangle$ s.t. there exists no $Conf''$ where $Conf \ominus Conf'' \subset Conf \ominus Conf'$

C4 If $S = A_{gen}(AT)$ then $A_{just}(AT, S) = AT$

C5 $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S)) = A_{gen}(A_{just}(AT, S))$

Property C1 articulates the principle of categorical matching which requires, in our instances that the result of an argumentation theory change should be an argumentation theory. Property C2 behaves like the AGM *success* postulate. Property C3 ensures that the conflict theory is minimally modified. Property C4 behaves in a similar fashion to the AGM *vacuity* postulate. Property C5 behaves in a similar fashion to the AGM *recovery* postulate.

Theorem 4.4.2. *A mixed-initiative argumentation system $\langle A_{gen}, A_{just} \rangle$, where A_{just} is an implementation of the conflict modifying decision justification procedure described in section 4.3.2 satisfies (C1) - (C5).*

Proof:

Given that A_{just} is an implementation of the conflict modifying decision justification procedure described in section 4.3.2, let us consider the following properties:

C1: $A_{just}(\langle AR, Conf, Pref \rangle, S)$ is an argumentation theory. The principle of categorical matching: Let us assume that A_{just} does not satisfy C1. This entails that A_{just} does not return an argumentation theory. Hence A_{just} is not a conflict modifying decision justification function. Thus, the assumption is violated.

C2: $S = A_{gen}(A_{just}(AT, S))$: Let us assume that A_{just} does not satisfy C2. This entails that A_{just} does not return $AT' = \langle AR, Conf', Pref \rangle$ such that S is the unique decision. Hence by definition (Definition 4.2.7), A_{just} is not conflict modifying decision justification function. Thus, the assumption is violated.

C3: $A_{just}(\langle AR, Conf, Pref \rangle, S) = \langle AR, Conf', Pref \rangle$ s.t. there exists no $Conf''$ where $Conf \ominus Conf'' \subset Conf \ominus Conf'$: Let us assume that A_{just} does not satisfy C3. This entails that there exists $Conf''$ where $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and S is the unique decision of $A_{just}(\langle AR, Conf'', Pref \rangle, S)$. This entails that there exists $\alpha_i \in AR$ and there does not exist any $\alpha_j \in S$ such that $conf(\alpha_j, \alpha_i) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf''$. It then follows that there exists $\alpha_i \in AR$ that is not defeated by any $\alpha_j \in S$. Hence, S is not an extension of $A_{just}(\langle AR, Conf'', Pref \rangle, S)$. Thus violating the assumption.

C4: If $S = A_{gen}(AT)$ then $A_{just}(AT, S) = AT$: Let us assume that A_{just} does not satisfy C4. This entails that A_{just} does not return AT. It then follows that A_{just} returns a modified $AT' = \langle AR, Conf', Pref \rangle$ where $Conf' \neq Conf$. However theorem 4.3.10 shows that A_{just} will perform minimal change to the AT. Since $S = A_{gen}(AT)$ then there is no change to AT. Hence, it must be the case that $Conf' = Conf$. Thus, the assumption is violated.

C5: $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S)) = A_{gen}(A_{just}(AT, S))$: Let us assume that A_{just} does not satisfy C5. This entails that $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S))$ does not return S. It then follows that A_{just} is not sound. However theorem 4.3.8 shows that A_{just} is sound. Thus, the assumption is violated. $\square$

Properties For Preference & Conflict Theory Change

Let us consider the set of rationality properties governing the preference & conflict modifying decision justification function. Given $AT = \langle AR, Conf, Pref \rangle$, a decision $S \subseteq AR$, a decision generation function A_{gen} and a decision justification function A_{just}. The rationality properties are:

PC1 $A_{just}(\langle AR, Conf, Pref \rangle, S)$ is an argumentation theory.

PC2 $S = A_{gen}(A_{just}(AT, S))$

PC3 $A_{just}(\langle AR, Conf, Pref \rangle, S) = \langle AR, Conf', Pref' \rangle$ s.t. there exists no $Conf''$ and $Pref''$ where $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and $Pref \ominus Pref'' \subset Pref \ominus Pref'$

PC4 If $S = A_{gen}(AT)$ then $A_{just}(AT, S) = AT$

PC5 $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S)) = A_{gen}(A_{just}(AT, S))$

Property PC1 articulates the principle of categorical matching which requires that the result of an argumentation theory change should be an argumentation theory. Property PC2 behaves like the AGM *success* postulate. Properties PC3 ensures that the conflict and preference theories respectively are minimally modified. Property PC4 behaves in a similar fashion to the AGM *vacuity* postulate. Property PC5 behaves in a similar fashion to the AGM *recovery* postulate.

Theorem 4.4.3. *A mixed-initiative argumentation system* $\langle A_{gen}, A_{just} \rangle$, *where* A_{just} *is an implementation of the **general decision justification procedure** described in section 4.3.4 satisfies (PC1) - (PC5).*

Proof:

Given that A_{just} is an implementation of the **general decision justification procedure** described in section 4.3.4, let us consider the following properties:

PC1: $A_{just}(\langle AR, Conf, Pref \rangle, S)$ is an argumentation theory. The principle of categorical matching: Let us assume that A_{just} does not satisfy PC1. This entails that A_{just} does not return an argumentation theory. Hence A_{just} is not a general decision justification function. Thus, the assumption is violated.

PC2: $S = A_{gen}(A_{just}(AT, S))$: Let us assume that A_{just} does not satisfy PC2. This entails that A_{just} does not return $AT' = \langle AR, Conf', Pref' \rangle$ such that S is the unique decision. Hence by definition (Definition 4.2.6 and Definition 4.2.7), A_{just} is not general decision justification function. Thus, the assumption is violated.

PC3: $A_{just}(\langle AR, Conf, Pref\rangle, S) = \langle AR, Conf', Pref'\rangle$ s.t. there exists no $Conf''$ and $Pref''$ where $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and $Pref \ominus Pref'' \subset Pref \ominus Pref'$: Let us assume that A_{just} does not satisfy PC3. This entails that there exists $Pref''$ and $Conf''$ such that $Pref \ominus Pref'' \subset Pref \ominus Pref'$, $Conf \ominus Conf'' \subset Conf \ominus Conf'$ and S is the unique decision of $A_{just}(\langle AR, Conf'', Pref''\rangle, S)$. This entails that there exists $\alpha_i \in AR$ and there does not exist any $\alpha_j \in S$ such that $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref'$ and $pref(\{\alpha_j\}, \{\alpha_i\}) \notin Pref''$ and $conf(\alpha_j, \alpha_i) \in Conf'$ and $conf(\alpha_j, \alpha_i) \notin Conf''$. It then follows that there exists $\alpha_i \in AR$ that is not defeated by any $\alpha_j \in S$. Hence, S is not an extension of $A_{just}(\langle AR, Conf'', Pref''\rangle, S)$. Thus violating the assumption.

PC4: If $S = A_{gen}(AT)$ then $A_{just}(AT, S) = AT$: Let us assume that A_{just} does not satisfy PC4. This entails that A_{just} does not return AT. It then follows that A_{just} returns a modified $AT' = \langle AR, Conf', Pref'\rangle$ where $Pref' \neq Pref$ or $Conf' \neq Conf$. However theorem 4.3.10 and theorem 4.3.4 show that A_{just} will perform minimal change to the AT. Since $S = A_{gen}(AT)$ then there is no change to AT. Hence, it must be the case that $Pref' = Pref$ and $Conf' = Conf$. Thus, the assumption is violated.

PC5: $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S)) = A_{gen}(A_{just}(AT, S))$: Let us assume that A_{just} does not satisfy PC5. This entails that $A_{gen}(A_{just}(A_{just}(A_{just}(AT, S), S'), S))$ does not return S. It then follows that A_{just} is not sound. However theorem 4.3.16 shows that A_{just} is sound. Thus, the assumption is violated. $\qquad\square$

Note that a recent work has been done on revision in argumentation systems in [35, 36] and Amgoud et al. [16]. Our approach differs from that of Cayrol et al. [35, 36] and Amgoud et al. [16]. Cayrol et al. addressing the issue of extension revision. Amgoud et al. considered the issue of revision of the argumentation theory without explicitly considering minimal changes. Both approaches do not capture preferences in the argumentation system nor the role preferences play in argumentation-base revision. Our approach focuses on the minimal revision of the argumentation-base rather then minimal revision of the generated extensions. Furthermore, our approach performs argumentation-base revision in a mixed-initiative fashion. Using mixed-initiative interaction to achieve the desired argumentation theory change allows for incremental changes from the previous argumentation-base, reflecting the essences of iterative revision. This approach provides us with the ability to perform traceability on decisions as well as retrospective reasoning in an argumentation framework.

In following section, we will present a discussion on the flexibility of the mixed-initiative argumentation framework.

4.4.2   Argumentation Theory Structures & Schemas

In this section, we will present a discussion on the flexibility of the mixed-initiative argumentation framework. The general focus will be on the structure (or lack of) in the arguments and background knowledge-base. Of interest is the combination of abstract arguments, structured arguments, the existence of a background knowledge-base and non existence of a background knowledge-base. There are four basic scenarios of interest:

- Abstract arguments and no background knowledge-base. In this setting, the underlying structure of the arguments is not defined. Most argumentation framework such as [22, 52] utilises such an approach. Furthermore, we assume the background knowledge-base for decision justification is empty. Essentially, starting from a clean slate.

- Structured arguments with no background knowledge-base. In this setting, some underlying structure is defined for the arguments. Given some underlying structure, different forms of attack such as rebuttal, assumption attack as well as direct and indirect attacks can be defined. This is typical for most argumentation systems such as [5, 8, 28, 34, 123, 124, 145, 147] where the interest is in the interplay between arguments. Again, in this setting, we assume the background knowledge-base for decision justification is empty.

- Abstract arguments with a background knowledge-base. In this setting, we assume no underlying structure for the arguments as proposed above. However, there exists some existing background knowledge-base. In such a situation, more considerations will need to be made in how the content of the knowledge-base is represented and revised.

- Structured arguments with a background knowledge-base. In this setting, there is some underlying structure for the arguments as proposed above. Furthermore, there exists some existing background knowledge-base. In such a situation, the combination of both structure in the arguments and the background knowledge-base increase the complexity of the decision justification. Part of the complexity is due to the fact that different forms of attack can now be specified and each form of attack may require different treatment during the decision justification process.

Abstract Arguments With No Knowledge-Base

Without defining an explicit structure to arguments, an argumentation schema and knowledge-based, a mixed-initiative argumentation system will reflect that of the framework proposed in the previous section (Section 4.2). We can view this as the most general form of mixed-initiative argumentation system. Arguments are represented as abstract object and interaction between arguments will be captured as attack and defeat relations. The different types of attack relations such as rebuttal, undercut or assumption attack are simply generalised to the attack relation. However, by using preferences to aid in identifying defeat (i.e. a successful attack), the notion of undercut as defined in [125] can also be captured. The use of preference to overwrite the attack relation provides user choice when applying rules. Without an explicit knowledge-based, the decision generation procedure is irrelevant. The first port of call for the mixed-initiative argumentation system is to accept any decision and hence the mixed-initiative interaction process provides the ability to subsume the provided knowledge. This allows the mixed-initiative argumentation system to evolve over a sequence of interactions, learning the background knowledge such as interactions between arguments as well as user preferences. The general machinery for mixed-initiative argumentation framework remains the same and no modification is required for both the decision generation and the decision justification procedures.

Abstract Arguments With Knowledge-Base

Without an explicit structure to arguments, argumentation schema, the arguments within the a mixed-initiative argumentation system will perform similarly to that describe in the previous section (Section 4.2). Arguments are represented as abstract object and interaction between arguments will be captured as attack and defeat relations. However, given an explicit knowledge-based, the mixed-initiative argumentation system now has a baseline to compare decisions. Using the decision generation component of the mixed-initiative argumentation system, the system is able to compare and validate decisions provided from a user interaction. During the validation process, should the decision differ, the decision justification process is triggered. This puts the system into "learning" or revision phase and allowing the mixed-initiative argumentation system to evolve over by subsuming the provided decision justifications. These justifications are used to enrich the background knowledge for future decision generation. The general machinery for mixed-initiative argumentation framework remains the same and no modification is required for both the decision generation and the decision justification procedures.

Structured Arguments With No Knowledge-Base

Let us assume a structured arguments with a premise-conclusion pair structure. Such a structure allows us to construct different forms of interaction between arguments such as defining a rebuttal, assumption attack or undercut. Other structures such as a chain or sequence of arguments further allow us to capture the notion of direct and indirect attacks. Similar to the above, without an explicit knowledge-based, the decision generation procedure is irrelevant. Hence the mixed-initiative interaction process provides the ability to subsume the provided knowledge. This allows the mixed-initiative argumentation system to evolve over a sequence of interactions, learning the background knowledge such as interactions between arguments as well as user preferences. The general machinery for mixed-initiative argumentation framework remains the same, however refine version of the decision generation and the decision justification procedures will need to be defined to handle the specific forms of argument interactions.

Structured Arguments With Knowledge-Base

Having an explicit argument structure and argumentation schema defined as well as an explicit background knowledge-base, the mixed-initiative argumentation system now has a baseline to compare decisions. Furthermore, this allows the system to capture different forms of interaction between arguments such as defining a rebuttal, assumption attack or undercut. Other structures such as a chain or sequence of arguments further allows the system to capture the notion of direct and indirect attacks. This constitutes the most knowledge rich scenario. Given all the different types of interactions, a more refine decision generation procedure will have to be defined. This decision generation procedure will now have to take into consideration the different notion of strength associated with each of the different forms of attacks. Furthermore, the revision operation will also have to be refined to consider the different forms of interplay. Although the general principle described in the previous section (Section 4.2) for decision generation and decision justification still holds, more specific procedures will have to be designed to address the added complexity as a result from the introduction of these different forms of interactions.

In following section, we will present a summary of the main concepts this discussed this chapter.

4.5 Summary

In this chapter, we have presented the mixed-initiative argumentation framework ($MIAF$). Within this chapter, we have motivated the need for a mixed-initiative argumentation framework where argumentation theory is revised due to a sequence of decisions. The proposed an abstract mixed-initiative argumentation framework used consists of two main components: decision generation and decision justification. The decision generation component consists of traditional argumentation of the style of Dung [52] and Bench-Capon [22]. The decision justification component takes as input a decision from the user and tips the generation machinery on its head and ask for a minimal revision the argumentation theory. Interleaving these two components results in a mixed-initiative argumentation framework that provides decision support and allows for the argumentation theory change over a sequence of decisions, hence, allowing for the management of decision justifications.

By utilising mixed-initiative interaction, the argumentation theory can be revised to reflect the changes in the "real-world". We have presented a collection of revision procedures for the modification of the preference theory and conflict theory within the argumentation theory. We have shown that the procedures are sound and terminates. We have also shown that these procedures results in a minimal change to the associated theories. Furthermore, we have proposed a set of properties governing the desirable behaviour for the argument theory change operators.

Finally, we presented a discussion on the flexibility of such a framework focusing on the combination presence/absence of structure in arguments and the presence/absence of a guiding background knowledge-base.

The following two chapters will illustrate the application of the two frameworks discussed in this part of the dissertation.

5

Applications of Argumentation Frameworks[1]

"The difficult part in an argument is not to defend one's opinion,
but rather to know it."
– Andre Maurois

5.1   Introduction

In chapter 3 and 4, we introduced the abstract preference-based accrual argumentation framework ($PAAF$) and the mixed-initiative argumentation framework ($MIAF$) respectively. In this chapter, we present two distinct applications of these frameworks.

In section 5.2, we will firstly introduce the notion of source-sensitivity in argumentation and

---

[1]Some work presented in this chapter also appeared in [37–40]

present a source-sensitive argumentation system. Within this section, we will firstly present a semi-ring based accrual argumentation framework ($PAAF_{SR}$). Secondly, we will present the source-sensitive argumentation system ($SSAS$) as an instance of the the semi-ring based accrual argumentation framework. We will also present a discussion on the applicability of this approach in areas such as defeasible argumentation and multi-agent negotiation. In section 5.3, we will introduce the medical group decision support tool, Just Clinical. Within this section, we will describe the implementation and evaluation of the tool. We will also discuss issues encountered in the use of the tool. In section 5.4, we present a summary of key ideas and concepts presented in this chapter.

5.2 Source-Sensitive Argumentation System

In chapter 3, we introduced the abstract preference-based accrual argumentation framework ($PAAF$). The recent applications of argumentation in multi-agent systems have drawn great interest. Systems such as [24, 119, 150] leverage formalisations of defeasible or non-monotonic reasoning using arguments. However, these forms of argumentation are insufficient in capturing many real-life instance of argumentation. Existing argumentation systems [52, 76, 86, 107, 113, 136, 145, 150] do not consider the source of the argument from the argument when evaluating defeat. In this section, we will focus on how the of source of an argument can play an important role in argumentation. We will demonstrate that arguments can be associated or labelled with the source of the arguments within the abstract preference-based accrual argumentation framework. The preference-based accrual argumentation framework is modified such that the source (agent) of an argument determines how preferred the argument is. We believe that when argumentation is utilised as a machinery for conflict resolution within a multi-agent system, the validity, acceptability and strength of an argument cannot be captured by the argument alone. It is common, in real life, for one to evaluate the strength and validity of arguments with respect to the provider of the argument. Associating agents to arguments highlight the ability of an argumentation system to capture the intuition of argument or information ownership. How an individual is perceived in one's community will influence the acceptability criteria and hence the force of one's arguments. Furthermore, such an association influences the applicability of argumentation rules as identified by Verheij in [147]. In [38], arguments are associated with their sources and the ranking of the arguments provides the notion of credibility. Such an association inevitably entails a subsequent change in the interpretation of the defeat relation and hence changes the dynamics of the argumentation system. The preference orderings on arguments can capture notions

such as degree of reliability, credibility or level of trust. We view argumentation as being (simultaneously) a process for information exchange, a process for conflict resolution and an approach to knowledge representation and reasoning. Multi-agent argumentation does not focus on logical "truth" alone but on convincing/persuading other agents of a particular view or position. In the rest of this section, we will explore how the credibility of the source of an argument can be brought to bear in the argumentation machinery in two distinct settings. In the first setting, we explore how credibility can be leveraged in an abstract argumentation framework (similar to those discussed in chapter 3). In the second setting, we consider arguments with internal structure, in particular, a structure consisting of the set of facts/premises, the set of assumptions and the set of conclusions.

In section 5.2.1, we will present an illustrating example to further highlight our viewpoint and approach. We will motivate the need for the sources to be associated with arguments when evaluating their strength and hence influencing the defeat relationship between these arguments. Furthermore, such an association influences the applicability of argumentation rules as identified by Verheij in [147]. The ordering of sources allows us to determine which source is more credible and hence which sets of rules are applicable. In section 5.2.2, we will firstly introduce the semi-ring based accrual argumentation framework ($PAAF_{SR}$). This framework is a modified version of the abstract preferences-based accrual argumentation framework ($PAAF$). Secondly, we will present the source-sensitive argumentation system ($SSAS$) as an instance of the the semi-ring based accrual argumentation framework. Utilising several examples, we will firstly illustrate the ability for $PAAF_{SR}$ to perform source-sensitive argumentation and secondly highlight some of the unique features of the source-sensitive argumentation system ($SSAS$). In section 5.2.3, we present a discussion on the applicability of this approach in areas such as defeasible argumentation and multi-agent negotiation.

5.2.1 Illustrating Example

In section 1.2.3, we illustrate the need to associate sources with arguments. The association of sources with arguments is utilised to evaluate their strength as well as the applicability of argumentation rules. In this example, we will further highlight the issue of accrual of arguments from multiple sources. Let us consider an extract of an example from Verheij [147]. The example starts with the planning of a picnic by John and Mary for Sunday. According to the national weather report, it is going to rain the whole day. Hence, they conclude that it will rain. We can hence construct an argument supporting the claim that it will rain:

National Weather Bureau: *It is going to rain on Sunday, therefore it will rain.*

Given that the argument is constructed based on information from the weather bureau (consisting of experts at meteorology), it is reasonably credible. Furthermore, there does not exist any other argument against such an argument. It is therefore reasonable to conclude that it will rain on Sunday. However, come Sunday, John observes that the sky is completely cloudless. Furthermore, this observation is concurred by Mary. We can therefore construct arguments supporting the claim that it will not rain:

John: *The sky is completely cloudless, therefore it will not rain.*
Mary: *The sky is completely cloudless, therefore it will not rain.*

Although both John and Mary are not experts at meteorology, we might be persuaded by the arguments by two credible people. Firstly, the arguments are constructed from observations representing the current state of affairs. Secondly, the arguments are constructed by two independent source of information. Note that in this situation, if sources are not associated with the arguments, we will not be able to distinguish between the arguments from John and Mary. As such, the utterance from John and Mary might be treated as one, resulting in the reduction of persuasive force of the argument in question.

Let us assume that John's father (a local farmer) weighs into the debate. John's father believes (but does not have definitive proof) that the national weather bureau is not very good at predicting local weather events, but says that he nevertheless thinks that it will rain. Since we don't have definitive proof which suggests that the national weather bureau is not good at predicting local weather events, we can only make such an assumption. From the above discussion, we can construct an argument supporting the claim that it will rain:

John's father: *Assumes that the national weather bureau is not good at predicting local weather events, it will still rain, therefore it will rain.*

Given John's father is a local farmer, he may have a more in-depth understanding of the local weather events. His credibility may outweigh that of John and Mary. Although his assumptions discredits the national weather bureau, his conclusion supports the conclusion of the national weather bureau. This poses several questions. Should arguments with conflicting premises but supporting conclusions be accrued? Would such a union of arguments strengthen or weaken the position? If we were to ignore the argument from the national weather bureau, is it a consequence of an attack on the argument or an inappropriate application of the argumentation rule or even the credibility of the source? Does the end justify the

means? This example highlights how the acceptability criteria of arguments in a dynamic and source-sensitive environment, coupled with the accrual of arguments requires specialised representation and reasoning machinery.

In this spirit, the next section will present the formalisation of the semi-ring based accrual argumentation framework ($PAAF_{SR}$) and the source-sensitive argumentation system ($SSAS$) as an instance of the the semi-ring based accrual argumentation framework to address the highlighted issues.

5.2.2   Formal System

In this section, we will explore how the credibility of the source of an argument can be brought to bear in the argumentation machinery in two distinct settings. In the first setting, we explore how credibility can be leveraged in an abstract argumentation framework (similar to those discussed in chapter 3). In the second setting, we consider arguments with internal structure, in particular, a structure consisting of the set of facts/premises, the set of assumptions and the set of conclusions.

Semi-ring-based Accrual Abstract Argumentation Framework

In the chapter 3, we have presented an abstract accrual argumentation framework that utilises a total ordering on the preferences to define a notion of defeat as well as accrual. This total ordering requirement maybe too strict. We will now present an alternate formulation utilising an algebraic approach to handle a set of abstract preference values that are partially ordered (as opposed to the earlier total order). The algebraic structure utilised is a semi-ring. A semi-ring structure consists of a set of abstract values, an additive operator and a multiplicative operator. The semi-ring additive operator allows us to perform comparison of the abstract values and the multiplicative operator performs a notional accrual of the abstract values.

Definition 5.2.1. *A semi-ring based accrual argumentation framework ($PAAF_{SR}$) is a triple:*

$$PAAF_{SR} = \langle AR, attacks_{SR}, Bel \rangle$$

where

- AR *is a set of arguments.*

- $attacks_{SR}$ *is a binary relations on* AR, *ie.* $attacks_{SR} \subseteq AR \times AR$.

- $Bel = \langle V, \oplus, \otimes, \Phi \rangle$ *is a semi-ring where*

    - V *is a set of abstract values.*

    - $\oplus$ *is an associative, commutative, idempotent and closed binary operator on* V.

    - $\otimes$ *is an associative, commutative and closed binary operator on* V.

    - Φ *is a total valuation function which maps elements of* AR *to elements of* V

    - $\otimes$ *distributes over* $\oplus$

The idempotent property on the $\oplus$ operator allows us to define a partial order $\leq_{pref}$ over the set of abstract values V. Such partial order is defined as: $\forall(v_1, v_2 \in V), v_1 \leq_{pref} v_2$ if and only if $v_1 \oplus v_2 = v_1$. Intuitively, $v_1 \leq_{pref} v_2$ means that v_1 is more preferred than v_2. For readability, we will denote $attacks(\alpha, \beta)$ to mean α attacks β and $pref(v_1, v_2)$ to mean $v_1 \leq_{pref} v_2$ or v_1 is preferred to v_2. Intuitively, the $\oplus$ operator allows us to perform comparisons between two semi-ring values while the $\otimes$ operator allows us to combine two semi-ring values. Note that we will use the symbols $\sum$ and $\prod$ to refer to the semi-ring operators $\oplus, \otimes$ in prefix notation.

With the exception of the notion of **preferable**, other notions such as conflict-freedom, defeat relation, maximal defeat relation, admissible set of arguments, accrued extension and preferred extension are defined the same as within $PAAF$. Due to the change in the valuation function and the introduction of the two semi-ring operators, the definition of preferable needs to be modified accordingly. The modified definition for **preferable** is therefore:

Definition 5.2.2. *Given a* $PAAF_{SR} = \langle AR, attacks_{SR}, Bel \rangle$ *and* $A, B \subseteq AR$ *be two sets of arguments.* A *is **preferable** to* B *(denoted as **preferable**(A, B)) if and only if:*

$$pref(\prod_{a_i \in A} \Phi(a_i), \prod_{b_i \in B} \Phi(b_i)).$$

The next two examples illustrates the use of $PAAF_{SR}$ in different situations. Let us consider example 3.1.1 in the $PAAF_{SR}$ framework.

Example 5.2.1. *Again, let us assume that* $\{\alpha, \beta, \gamma\}$ *represents* {*"Tom has not been jogging for several days, so he should go jogging", "It is raining, so Tom should not go jogging", "It is hot, so Tom should not go jogging"*} *respectively. We will also assume the existence of a set of abstract values* $\{v_1, v_2, v_3, v_4\}$ *with the ordering of* $v_1 \leq v_2, v_2 \leq v_3, v_3 \leq v_4$.

Let the valuation function Φ maps the following arguments to their respective preference values and the two binary operators $\oplus, \otimes$ on the abstract values V be defined extensionally as shown in table 5.1 (as with similar tables presented in chapter 3, the table summarises other dimensions of the problem as well).

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{SR}$	$attacks_{SR}(\alpha, \gamma), attacks_{SR}(\gamma, \alpha), attacks_{SR}(\beta, \gamma), attacks_{SR}(\gamma, \beta)$
V	$\{v_1, v_2, v_3, v_4\}$
Φ	$\Phi(\alpha) = v_3, \Phi(\beta) = v_2, \Phi(\gamma) = v_2, \Phi(\emptyset) = v_4$
$\oplus$	$v_1 \oplus v_2 = v_1, v_1 \oplus v_3 = v_1, v_1 \oplus v_4 = v_1, v_2 \oplus v_3 = v_2, v_2 \oplus v_4 = v_2, v_3 \oplus v_4 = v_3$
$\otimes$	$v_1 \otimes v_1 = v_2, v_2 \otimes v_2 = v_4, v_3 \otimes v_3 = v_4, v_4 \otimes v_4 = v_4,$ $v_1 \otimes v_2 = v_3, v_1 \otimes v_3 = v_4, v_1 \otimes v_4 = v_4, v_2 \otimes v_3 = v_4, v_2 \otimes v_4 = v_4, v_3 \otimes v_4 = v_4$
admissible	$\emptyset, \{\alpha\}$
accrued	$\{\alpha\}$
preferred	$\{\alpha\}$

This table provides a summary of the discussion in example 5.2.1 and illustrates the computation of the admissible sets of arguments, accrued extension and preferred extension.

Table 5.1: Summary of Example 5.2.1

From this example, we can show that when arguments β or γ are presented separately, γ defeats α and β defeats α. However, when both β and γ are presented together, the accrual causes α to be undefeated.

Let us now consider example 1.2.2 in the $PAAF_{SR}$ framework.

Example 5.2.2. *Let us assume that $\{\alpha, \beta, \gamma\}$ represents $\{$"Bill has robbed someone, so he should be jailed.", "Bill has assaulted someone, so he should be jailed.", "Bill is a juvenile, therefore he should not go to jail."$\}$ respectively. We will also assume the existence of a set of abstract values $\{v_1, v_2, v_3, v_4\}$ with the ordering of $v_1 \le v_2, v_2 \le v_3, v_3 \le v_4$. As with previous example, the details are summarised in the table 5.2 .*

From this example, we can show that when arguments α or β are presented separately, γ defeats α and γ defeats β. However, when both α and β are presented together, the accrual causes γ to be defeated.

There exist several differences between the formulations of $PAAF$ and $PAAF_{SR}$. The differences are as follows:

- The valuation function Φ in the $PAAF$ maps $2^{AR} \longrightarrow V$ where as in $PAAF_{SR}$, it is a mapping from $AR \longrightarrow V$.

AR	$\{\alpha, \beta, \gamma\}$
$attacks_{SR}$	$attacks_{SR}(\alpha, \gamma), attacks_{SR}(\gamma, \alpha), attacks_{SR}(\beta, \gamma), attacks_{SR}(\gamma, \beta)$
V	$\{v_1, v_2, v_3, v_4\}$
Φ	$\Phi(\alpha) = v_3, \Phi(\beta) = v_3, \Phi(\gamma) = v_2, \Phi(\emptyset) = v_4$
$\oplus$	$v_1 \oplus v_2 = v_1, v_1 \oplus v_3 = v_1, v_1 \oplus v_4 = v_1, v_2 \oplus v_3 = v_2, v_2 \oplus v_4 = v_2, v_3 \oplus v_4 = v_3$
$\otimes$	$v_1 \otimes v_1 = v_1, v_2 \otimes v_2 = v_1, v_3 \otimes v_3 = v_1, v_4 \otimes v_4 = v_1,$ $v_1 \otimes v_2 = v_1, v_1 \otimes v_3 = v_2, v_1 \otimes v_4 = v_3, v_2 \otimes v_3 = v_1, v_2 \otimes v_4 = v_2, v_3 \otimes v_4 = v_1$
admissible	$\emptyset, \{\alpha, \beta\}$
accrued	$\{\alpha, \beta\}$
preferred	$\{\alpha, \beta\}$

This table provides a summary of the discussion in example 5.2.2 and illustrates the computation of the admissible sets of arguments, accrued extension and preferred extension.

Table 5.2: Summary of Example 5.2.2

- $PAAF_{SR}$ provides two semi-ring operators to perform comparisons and combinations.

- The preference values in $PAAF$ is totally ordered where as it need not be in $PAAF_{SR}$.

Source Sensitive Argumentation System

In this section, we will provide an outline of a source sensitive argumentation system based on work from [38]. We will consider arguments with internal structure consisting of: the set of facts, the set of assumptions and the set of conclusions. For simplicity, we will take any finitely generated propositional language $\mathcal{L}$ with the usual punctuation signs, and logical connectives $\neg$ (not) and $\supset$ (implies). For any set of wffs $S \subseteq \mathcal{L}$ and any $\alpha \in \mathcal{L}, S \vdash \alpha$ means α is provable from premises S. For any set of wffs $S \subseteq \mathcal{L}, Cn(S) = \{\alpha | S \vdash \alpha\}$.

Definition 5.2.3. *(Argument) An argument α is a triple $\langle F, A, C \rangle$ where F, A and C denote the sets of facts, assumptions and conclusions respectively.*

Note that the formulation of arguments into facts, assumptions and conclusions corresponds to the formulations of several frameworks for abduction [79,80,118]. With such an approach, assumptions can now be explicitly asserted. Logically, the conclusion is the consequence of the union of facts and assumptions. There are several conditions that should be considered. Firstly, an assertion should not simultaneously be asserted as a fact and an assumption. Having an assertion being a fact and an assumption could cause inconsistency as notionally, assumptions are weaker than fact.

Definition 5.2.4. *(Well-founded Argument) An argument α is a* well-founded *argument if and only if it satisfies the following conditions:*

- $F, A, C \subseteq \mathcal{L}$

- $F \cap A = \emptyset$

- $F \cup A \vdash C$

- $F \cup A \cup C \nvdash \bot$

We will write F_α, A_α and C_α to respectively denote the facts, assumptions and conclusions associated with an argument α. One can view a well-founded argument as a *premise-conclusion* pair where the *premise* is $\langle F,\ A \rangle$ and the *conclusion* is C. By allowing assumptions, we permit a weaker notion of a premise for an argument. Note that since we are interested in *rational* agents, we have eliminated self-defeating [113,125] arguments. Notionally, assumptions are weaker than fact, but the strength of conclusions should be dependent on the premises it is based. There are many approaches to computing this notion of strength. Two boundary notions can be considered: sceptical and credulous. A sceptical reasoner may consider the strength of the conclusion as equivalent to the weakest element in the premise while a credulous reasoner may consider the strength of the conclusion as equivalent to the strongest element, hence, such an approach provides the spectrum of possibilities.

Definition 5.2.5. *(Conflict) A pair of well-founded arguments α and β are said to be in conflict if and only if* $(F_\alpha \cup A_\alpha \cup C_\alpha \cup F_\beta \cup A_\beta \cup C_\beta) \vdash \bot$

Each component of an argument (facts, assumptions and conclusions) have an implicit notion of strength. Given two well-founded arguments α and β, the basic interaction between components of an argument are as listed:

- Fact Conflict: $F_\alpha \cup F_\beta \vdash \bot$

- Assumption Conflict: $A_\alpha \cup A_\beta \vdash \bot$

- Conclusion Conflict: $C_\alpha \cup C_\beta \vdash \bot$

- Fact-Conclusion Attack: $F_\alpha \cup C_\beta \vdash \bot$

- Conclusion-Fact Attack: $C_\alpha \cup F_\beta \vdash \bot$

- Assumption-Conclusion Attack: $A_\alpha \cup F_\beta \vdash \bot$

- Fact-Assumption Attack: $F_\alpha \cup A_\beta \vdash \bot$

- Assumption-Fact Attack: $A_\alpha \cup F_\beta \vdash \bot$

Note that complex interaction/relations may consist of all combinations of the above. Note also that the above interactions can be separated into two classes: conflict where the interaction is symmetrical or omni-directional; attack where the interaction is uni-directional. This provides a vocabulary for describing all the different forms of interplay between arguments.

Definition 5.2.6. *(Semi-ring) A semi-ring is a triple* $\mathcal{G} = \langle V, \oplus, \otimes \rangle$ *where*

- V *is a set of abstract values.*

- $\oplus$ *is an associative, commutative, idempotent and closed binary operator on* V.

- $\otimes$ *is an associative, commutative and closed binary operator on* V.

- $\otimes$ *distributes over* $\oplus$

Each component of an argument $\langle F, A, C \rangle$ could be mapped to an abstract value and utilising the two semi-ring operators, a derived notion of argument strength can be computed. For instance, for any given argument α and a semi-ring $\mathcal{G}$ with $v_1, v_2 \in V$, if F_α is assigned v_1 and A_α is assigned v_2, we could potentially compute a value for C_α by performing $v_1 \otimes v_2$. Such an approach would capture the intuition that the strength of the conclusion is derived from the combination of the strength of the facts and assumptions. Furthermore, the strength of the argument can also be derived by using a similar approach. This interesting approach will fuel debate on the appropriateness of the strength assignment and the relative weighting on facts, assumptions and conclusions as found in studies in social choice theory such as [20,103,140]. Hence, it diverges from the aim of this chapter and it is outside the scope of this discussion. We have provided the vocabulary and machinery to allow exploration into these issues and will leave the details to the designer of the system. For the purpose of this discussion, we will focus on assignment of credibility to sources and arguments.

The notion of tagged arguments allow us to uniquely identify argument source. By tagging the arguments, we are simply labelling the arguments with additional information.

Definition 5.2.7. *(*Tagged *Arguments) Given a set of unique identifiers* $\mathcal{I}$, *we define* $\mathcal{A}$ *as a set of* tagged *arguments of the form* $\langle S, A \rangle$ *where*

- $S \in \mathcal{I}$ *represents the* tagged *arguments' source.*

- A *is a set of* well-founded *arguments.*

We will write S_ϕ and A_ϕ to respectively denote the source and well-founded arguments associated with a tagged argument ϕ.

Definition 5.2.8. *(Credibility Function) Given a set of unique identifiers $\mathcal{I}$ and a semi-ring $\mathcal{G} = \langle V, \oplus, \otimes \rangle$, we say Φ is a credibility function if it maps all values of $\mathcal{I}$ into V.*

The notion of credibility provides the agent with a measure of strength per source. For simplicity, we have utilised the valuation function Φ that maps a set of unique identifiers into a set of abstract values V. However, one could define an arbitrarily complex valuation function taking into account additional insight such as the context. For example, if the agent is a stock-broker, then any arguments related to the share market from this agent will be more credible than any agent that is not a stock-broker. We can further enrich the approach by define the use of a class of valuation function. This approach allows the system to be applicable to multiple context.

Definition 5.2.9. *(Source Sensitive Argumentation System) A source sensitive argumentation system is defined as:*

$$\mathcal{SAS} = \langle \mathcal{A}, \Phi, \mathcal{G} \rangle$$

where

- *$\mathcal{A}$ is a set of tagged arguments.*

- *Φ is a credibility function.*

- *$\mathcal{G}$ is a semi-ring.*

In the following section, we will present a brief discussion on the resulting consequence from the proposal of $PAAF_{SR}$ and $SSAS$.

5.2.3 Discussion

In the previous section, we have presented the semi-ring based accrual argumentation framework ($PAAF_{SR}$) and the source-sensitive argumentation system ($SSAS$). In this section, we will present a brief discussion on the impact of $PAAF_{SR}$ and $SSAS$ on existing argumentation systems.

The contribution of the $PAAF_{SR}$ and $SSAS$ is two fold. Firstly, by introduction the semi-ring and using the associated semi-ring operator, the abstract value for a set of accrued arguments can now be computed. The differentiation between $PAAF$ and $PAAF_{SR}$ is significant as in $PAAF$, the abstract value for a set of accrued arguments has to be explicitly asserted. This

provides the basic machinery to compute strength or credibility for a set of accrued arguments. Secondly, the introduction of structure to arguments. The value proposition for such an approach is to provide a vocabulary for describing different forms of interaction between components of the argument. We have proposed explicit recording of assumptions as it increases the expressiveness of an argument. By explicitly capturing assumptions, new forms of argument to argument interactions can be considered. It opens up the debate regarding acceptability of arguments with conflicting assumptions but non-conflicting facts and conclusions. The semi-ring also plays an important role. It provides the machinery to compute derived values for the conclusion. Thus, addressing a potential limitation of $PAAF$ where preference values have to be assigned to the power set of arguments in advance. These two modifications increase the expressiveness of the original framework and allows for addressing more complex problems.

We will consider two areas of applications: defeasible argumentation and argumentation-based negotiation.

Defeasible Argumentation

Although the works such as [24, 52, 53, 119, 150] focuses on representation and logical reasoning, we have shown that by introducing the notion of structure in arguments as well as argument sources, we are able to provide the vocabulary and machinery for a practical system. We have modelled our system around these collection of defeasible argumentation systems. We feel that our work will complement the advances already achieved. The direct applicability of the source-sensitive argumentation system ($SSAS$) and the semi-ring based accrual argumentation framework ($PAAF_{SR}$) to framework such as [24, 52, 53] is due to the fact that the original abstract preference-based accrual argumentation framework ($PAAF$) is based on these frameworks. In chapter 3.2 We have shown that there is a direct relation between $PAAF$ and AF proposed in [52]. Futhermore, we have shown that the source-sensitive argumentation system ($SSAS$) is an incremental refinement of the $PAAF_{SR}$.

Additionally, the proposed the source-sensitive argumentation system ($SSAS$) provides a rich vocabulary for expressing arguments and machinery capable of addressing issues highlighted in recent studies such as [28, 33, 121] in defeasible argumentation. In [121], Prakken focus on the fairness and soundness of dynamic argumentation protocol. In [33], Carbogim addressed issues associated to change in the underlying knowledge base caused by new arguments. In [28], Brewka dealt with meta-level argumentation, providing undercut via a notion of preference on which defeat rule holds. Components of these work have similarities to our

proposal.

Argumentation-Based Negotiation

The exchange of arguments and counterarguments has also been studied in the context of multi-agent interaction. In [107], Parsons et al. considered argumentation as a component of negotiation protocols, where arguments for an offer should persuade the other party to accept the offer. Generally, these works focus on the negotiation protocol and agent behaviours, leaving the representation and internal reasoning to the designer. In [86], Kraus et al. prescribed a range of agent behaviours. We feel that our work further enhance the internal reasoning of these agents. Although some notion of sources and credibility may have existed in studies such as [76, 136], these notions were not explicitly utilised when evaluating arguments. The vocabulary and machinery presented in the source-sensitive argumentation system ($SSAS$) can be deployed to complement these proposals.

The next section will present Just-Clinical, a group decision support tool based on the mixed-initiative argumentation framework.

5.3   Just-Clinical

In chapter 3 and 4, we introduced the abstract preference-based accrual argumentation framework ($PAAF$) and the mixed-initiative argumentation framework ($MIAF$) respectively. In this section, our sole aim is to present a proof of concept tool utilising the theoretical constructs from both $PAAF$ and $MIAF$. We will present a multi-disciplinary medical group decision support tool, *Just-Clinical*. *Just-Clinical* is a tool utilised for decision support in the treatment of cancer. It is important to note that group decision support is one of many possible applications of the abstract preference-based accrual argumentation and the mixed-initiative argumentation frameworks. One of the issues in medical decisions is the inconsistency in the application of rules for treatment. This is partially due to treatment decisions being recorded without substantiative justification for the decision. Such inaccuracies place the quality of the decisions into question. The quality assurance of decisions entails the retention of decision justifications such that retrospective analysis can be performed on decisions. Furthermore, using these past decisions and justifications, the "correctness" of the current decision with respects to this collection of past decisions are assessed. Hence, notionally maintaining a level of quality such that future decisions can be compared against.

In section 5.3.1, we present a multi-disciplinary (MDT) medical group decision session highlighting that multi-disciplinary medical group decision is a prime candidate for the use of argumentation and particularly the applicability of $PAAF$ and $MIAF$. In section 5.3.2, we will introduce the medical group decision support tool, Just Clinical. Within this section, we will describe the implementation and evaluation of the tool. In section 5.3.3, we will discuss issues encountered in the use of the tool.

5.3.1 Illustrating Example

Let us look at an extract from a multi-disciplinary medical group decision session taken from [40]. The discussion is on a patient with early stage superficial unilateral larynx cancer. The discussion involves several medical specialists (Surgeons S_1,S_2,S_3, Radiation Oncologists RT_1,RT_2) debating on the best treatment for the disease (Figure 5.1, 5.2 and 5.3). In general, the patient's physician will have the final say. In this scenario, although the physician did not partake nor was privileged to the discussion, the ultimate decision still lies with him/her.

Disease Definition: Larynx Cancer Early Superficial Unilateral	
S_1	: (A_1) My opinion is to take out the patient's larynx. This has the best cure rate of 99%.
S_2	: (A_2) I agree, taking out the patient's larynx would provide the best cure potential.
S_3	: (A_3) I also agree, taking out the patient's larynx would provide the best cure potential.
RT_1	: (A_4) But if taking out the patient's larynx, the patient will have no voice.
RT_1	: (A_5) However if you use radiotherapy, there is a 97% cure rate from the radiotherapy and about 97% voice quality, which is very good. The 3% who fail radiotherapy can have their larynx removed and most of these will be cured too.

Figure 5.1: Multi-disciplinary Clinical Decision Support 1: Larynx Cancer

In Figure 5.1, the arguments A_1 ... A_5 illustrates several important issues. Firstly, the need for accrual in argumentation. Within argumentation, "accrual" generally refers to the grouping of arguments to support or refute a particular opinion. It is recognised [92, 122, 146] that "accrual" of arguments is an issue that requires attention. To highlight our point, let us focus on three key arguments. A_4 forms the basis of an attack on the argument A_1. When just considering these two arguments alone, it maybe difficult to determine which course of action is the most appropriate. Now, let us consider the argument: A_1 in conjunction with the argument A_5. Again, it maybe difficult to determine which choice is a more appropriate action to take. However, when we consider all three arguments together, it is clear that the best course of action is to perform radiotherapy before taking out the patient's larynx. Secondly, the ability to strengthen arguments by repetition. To highlight our point, let us focus on the

arguments: A_1, A_2, A_3. Although these three arguments do not enlighten the discussion with any additional information, it is conceivable that in a human debate situation, the number of arguments is sufficient enough to overwhelm any suggestion of the contrary. However, we are not advocating that we should always strengthen a position simply by providing multitude of identical arguments. Performing such tasks should be informed by some additional information such as source's expertise or credibility. Finally, the importance of the information sources during argumentation. If we consider the accrual of identical arguments as a reflection of the norms of a community, then it is conceivable that the first course of action would be to take out the patient's larynx. However, if the specialist RT_1 has special insight or knowledge not shared with the other specialists (e.g., the specialist is the ONLY radiation oncologist in the group), therefore, might occupy a somewhat privileged position, it is then possible that the arguments made by this particular specialist may carry more weight. In this example, we motivate that the credibility of the individual presenting the argument is important. Using this notion of credibility, we can infer a preference ordering on the arguments.

S_2	: (A_6) My opinion is also that the patient should have a hemi-laryngectomy. This will give a cure rate is as good as radiation therapy.
S_3	: (A_7) I agree, performing a hemi-laryngectomy would give a cure rate as good as radiotherapy.
RT_1	: (A_8) Yes, I have performed many hemi-laryngectomies, and when I reviewed my case load, the cure rate was 97%, which is as good as that reported internationally for radiotherapy.
RT_2	: (A_9) I agree, however you fail to take into account the patient's age. Given the patient is over 75, operating on the patient is not advisable as the patient may not recover from an operation.
RT_1	: (A_{10}) Yes, however in this case, the patient's performance status is extremely good, the patient will most likely recover from an operation. (i.e. the general rule does not apply)

Figure 5.2: Multi-disciplinary Clinical Decision Support 2: Larynx Cancer

In Figure 5.2, arguments $A_6 \ldots A_{10}$ illustrate an interesting phenomenon. In this particular instance, the specialist RT_1 did not disagree with the correctness of the presented facts and the conclusion in the argument presented by RT_2, but rather the applicability of the underlying inference rule that is used to construct the argument. This phenomenon is defined by [110, 125] as "undercut". In this situation, the argument presented by RT_1 is more specific. This indicates that there exist some exceptions to the general decision rules that are context dependent and a revision on the attack relation is required.

In Figure 5.3, arguments A_{11} and A_{12} illustrate an attack on the user preference. Similar to the previous example, attacks on the user preference are generally context sensitive and may indicate a revision on the general attack relation. These two examples illustrate that an argumentation system should evolve over time, accumulating past decisions as justifications

S_2	: (A_{11}) Reviewing our past case decisions, evidence suggest that the we have always performed a hemi-laryngectomy, hence my preference is to do the same.
S_3	: (A_{12}) I agree, however, there is some new medical literature reporting that the voice quality after a hemi-laryngectomy was only 50% acceptable and the reporting institution was the North American leaders in hemi-laryngectomy, hence we should perform radiotherapy.

Figure 5.3: Multi-disciplinary Clinical Decision Support 3: Larynx Cancer

for future decisions. However, it is also clear that in some instances, we wish to overrule past precedent. In most argumentation and decision support systems presented in the literature, the systems are relatively static. Most systems are open to new facts, however have difficulties handling changing rules and preferences.

Furthermore, let us now assume that the patient's physician decided to perform a hemi-laryngectomy. He/She will now have to justify the decision. If we assume that the above discussion did not occur (i.e. empty knowledge base), then the physician only requires to present arguments for the hemi-laryngectomy decision. However, if the knowledge base consists of the arguments, attack relations and preferences captured from the discussion, then the physician will be required to not only present arguments for the hemi-laryngectomy decision but also address all attacks on his/her decision. One can view the sequence of interactions captured in the discussion as "decision generation" mode, if all the arguments, attack relation and preferences exist in the knowledge base (in other words, the knowledge base is complete) and we request the argumentation system to present us with a decision. Alternatively, if the knowledge base is incomplete, erroneous or an undesired decision generated, a decision can be introduced and modification performed on the knowledge base in the "decision justification" mode.

In spite of these shortcomings, these examples reinforce the view that argumentation is a prime candidate for such a group decision support situation. In the next section, we will present the medical group decision support tool.

5.3.2 Medical Group Decision Support Tool

In this section, we will introduce the medical group decision support tool, Just-Clinical. We will firstly provide a description of the problem, highlighting the various aspects and complexity of the problem domain. Secondly, we will describe the implementation, highlighting the unique philosophy and approach that was taken. Finally, we will present the use and evaluation of the tool.

Problem Description

In the view of improving patient care, it is common practise to have multi-disciplinary team (MDT) for cancer treatment. A multi-disciplinary team usually consists of individuals from a range of specialisation such as surgeons, radiation oncologist, medical oncologist, histopathologists and specialist nurses. The multi-disciplinary team is responsible for making key decisions with regard to the cancer treatment of the patient. Some of the key decisions to be made are in relation to survival/prognosis, treatment options, treatment side-effects and impact on quality of life. Decisions are also dependent on external factors such as psychological impact to patient and family, family history, availability of clinical trials, accessibility to treatment. This highlights that the treatment of cancer for any individual is a complex process.

We will briefly describe the different aspects of the problem that is taken into considerations. Firstly, a cancer can be described by three key pieces of information.

- Site of the disease

- Histology of the disease

- Staging of the disease

Site of the disease describes the location of the disease. This is usually presented as an ICD code indicating which part of the body the disease resides. The histology of the disease describes the usual progression of the disease based on pass data. This is usually presented as an ICD03 code. The staging of the disease describes which stage of disease is at. In our case, it is represented using TNM (Tumour, Nodes, Metastasis) coding. Tumour coding indicates size and growth stage of the solid tumour. Nodes coding indicates the involvements of lymph nodes and the extent of the involvement. Metastasis coding indicates that the disease has metastasised and the disease has spread to other organs or regions of the body. Given the three key pieces of information, we can uniquely identify a cancer. These three items of information combined also provide the basic prognosis of a patient having the disease.

The effect and side-effect of a treatment can be considered by the following:

- Survival

- Control

- Physical Toxicity

- Psychological Toxicity

- Clinician's choice

Survival describes the rate of survival for a patient given a particular treatment or therapy. This value is usually expressed as a percentage or probability of survival over a duration of 3, 5 years. Control describes the cancer response for a particular treatment and the ability for the treatment to control the growth and spread of the disease. Physical and psychological toxicity describe the affects the treatment have on the body. These assess the treatment side-effects and are usually graded between 0 to 5 over several criteria. Clinician's choice describes the clinician's choice of treatment based on his/her collected past experience in treating the particular type of cancer. The weighted sum of the combination of the 5 aspects provides an acceptability value for a given treatment.

There are three basic modalities of the treatment or therapy. These modalities are: radiation therapy, chemo therapy and surgery. However, it is rarely the case that only one modality is deployed in a treatment. The combination of modalities that can be considered are:

- Concurrent radiation and chemo therapy

- Radiation therapy salvaged by surgery

- Chemo therapy follow by radiation therapy salvaged by surgery

- Concurrent radiation and chemo therapy salvaged by surgery

- Surgery follow by radiation therapy

- Do nothing

The notion of "concurrent" basically expresses that multiple modalities are utilised at the sane time in treating a patient. The notion of "follow by" expresses that one modality of treatment is immediately followed by another modality of treatment. The notion of "salvaged by" expresses that the first modality is the main source of treatment and should that source of treatment fail, the subsequent modality is utilised as the next resort. Usually, an evaluation is performed after completion of the main treatment modality to determine if the treatment is having an affect on the disease or not. The last choice is to not perform any treatment. Although not strictly a treatment modality, this choice is used to monitor the growth of the

disease. It is important that this choice is available as in some cases the treatment side-effects are worst than having the disease.

Evidence supporting the treatment falls into the following three categories (ordered from strongest to the weakest):

- Level A: randomised controlled trial/meta-analysis

- Level B: other evidence

- Level C: consensus/expert opinion

Types of attack within each aspect are associated with the strength of the attack (lower numbers are stronger). Treatment analyses are performed over 5 categories: survival, control, physical toxicity, psychological toxicity and clinicians choice. These categories are addressed in stages.

Tool Implementation

In [141], Tecuci et al. proposed seven issues of mixed-initiative reasoning for consideration:

1. Task

2. Control

3. Awareness

4. Communication

5. Personalisation

6. Architecture

7. Evaluation

The task issue focuses on the need for the division of labour and responsibility between the human and the agent One of the dimension to consider is the complementarity abilities between a human and an automated agent. Considerations have to be made in relation to the difference in reasoning styles and computational strengths. The control issue focuses on the strategies to deploy when shifting the initiative and control the human and the agent. These

strategies dictate the timing of the interaction and interruption which may, in some cases be perceived as proactive behaviour. The awareness issue focuses on the maintenance of a shared understanding between the human and the agent in relation to the evolving state of the problem and environment. The communication issue focuses on the protocols that facilitate the exchange of knowledge and information between the human and the agent. These protocols include mixed-initiative dialogue and multi-modal interfaces. The personalisation issue focuses on the adaptability of the agent's knowledge and behaviour to suit the user's problem solving strategies, preferences, biases and assumptions. The architecture issue focuses on the design principles, methodologies, and technologies for different types of mixed-initiative roles and behaviours. The evaluation issue focuses on the human and automated agent contribution to the emergent behaviour of the system. Performance indicators such as fully automated, fully manual, or alternative mixed-initiative approaches are considered.

We have utilised these seven aspect as guiding principles in the construction of Just-Clinical. Utilising a Web 2.0 philosophy, we have constructed a web enabled medical group decision support system using Asynchronous JavaScript and XML (AJAX) with a back-end repository. HyperText Markup Language (HTML) and Javascript are used to build the user interface and controls the interaction with the web server. Hypertext Preprocessor (PHP) is used to build the reasoning engine to perform back-end computation of the arguments. MySQL is used as the database repository. The generation of decisions consists of a set of set operations performed in SQL. These operations are directly derived from the definition for an extension provided in chapter 3. The decision justifications procedure A.1, A.2, A.3 and A.4 can be found in in the appendix. These procedures describes a series of set operations. These operations are performed on the database utlised SQL. The benefits of this approach are platform independence, portability, scalability and accessibility. Note that this approach is distinctly different to that taken in the construction of Dungine [139] where the computation of extensions is achieved by using argument games.

Tool Evaluation

The prototype was presented to several oncologists and a "head-and-neck" session was simulated. A "head-and-neck" session is where groups of oncologists meet to discuss treatment therapy for cancer cases in the head to neck region. During this session, two typical larynx cancer cases were discussed. Treatment analyses are performed over 5 different aspects or categories. These categories are as listed: survival, control, physical toxicity, psychological toxicity and clinician's choice. These categories are addressed in stages. Argumentation is

performed at each stage and final recommendation is based on the accrual of all arguments over all the stages. Each stage can be viewed as a decision-making cycle where decision made affects the available choices for the next cycle. A summary of the two cases that were discussed can be found in table B.1 and B.2 (in the appendix). Each of the table presents the case description, decision variables and values. Given a case description, the system presents a possible recommendation (if one exists). A case is described by disease site, disease histology and disease staging information. A recommendation is presented as a treatment modality. Specialists are then asked if the recommendation is acceptable. If the recommendation is not acceptable, the system asks the specialist to select a recommendation and justify it with arguments, with which the system then recomputes a new recommendation. Justifications are expressed as statements and attack relations associated with an integer indicating the relative strength of the statement. This relative strength is computed (by the user) based on the type of evidence. There is a general consensus in the medical community as to ordering of the type of evidence, hence the computation of the strength is generally acknowledged as given. If the recommendation does not coincide, the system presents its findings and asks for more justifications. This process is iterated until the recommendation of the system coincides with the specialist's choice.

Figure 5.4: Treatment Choices

In Figure 5.4, we present the user interface. In the left column, pertinent details of the case definition are presented. In the right column, the users are presented with a list of possible treatment recommendations appropriate for the case profile. These treatment recommendations are extensions. In Figure 5.5, we present the argument modification interface. The
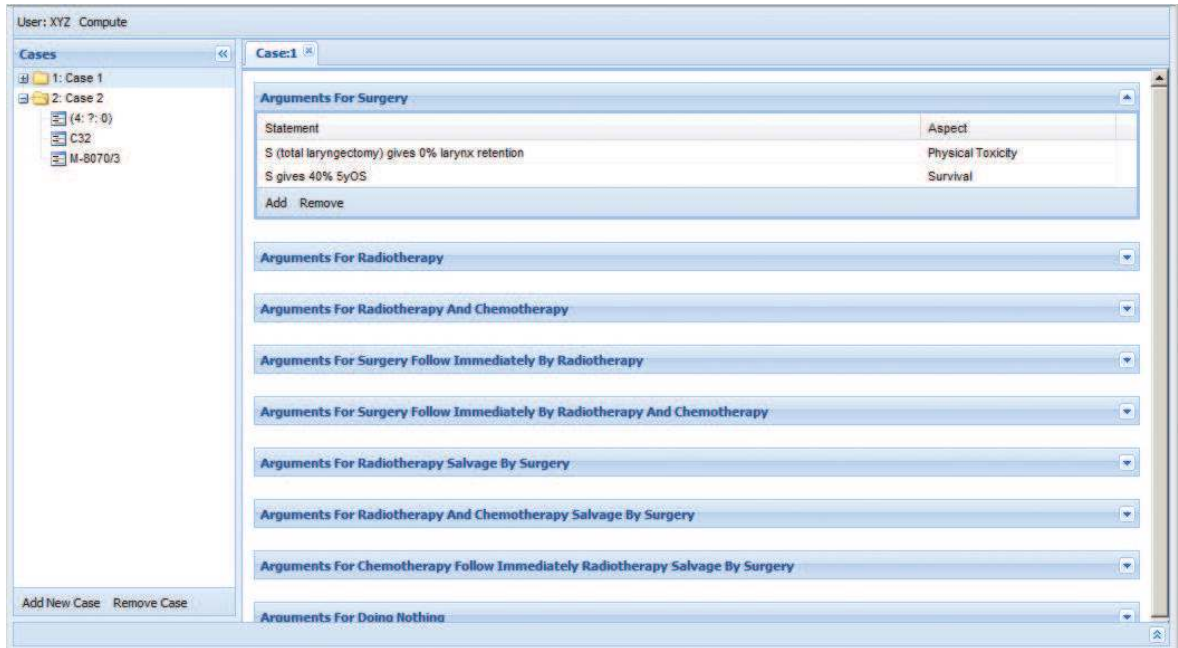
Figure 5.5: Arguments

users are allowed to add, delete and modify the arguments associated with a particular treatment choice in the forward learning mode. The changes require that the clinician provide the strength of the evidence and a literature references if used. In essence, by associating the argument with the treatment choice, the user has provided justification for the particular treatment. In Figure 5.6, we present the resulting output, which illustrates the recommended decision for each facet of a given sequence of decisions. Each facet has different priority (if two treatments have identical cure and control rates, the one with lower physical toxicity is preferred) and the final treatment choice is computed using these preferences. This figure also illustrates the ability for the user to validate the recommendations and subsequently activate the second of the two learning modes where the user disagrees with the recommendation.

In the next section, we will discuss issues encountered in the use of the tool.

5.3.3 Discussion

The general response to using the tool has been positive. The specialists found that the tool is useful both as a practical tool for a trained specialist as well as a teaching aid for medical registrar. The specialists also felt that the tool has huge potential, however, since the tool is still in its infancy there are still several practical issues that need addressing. During the
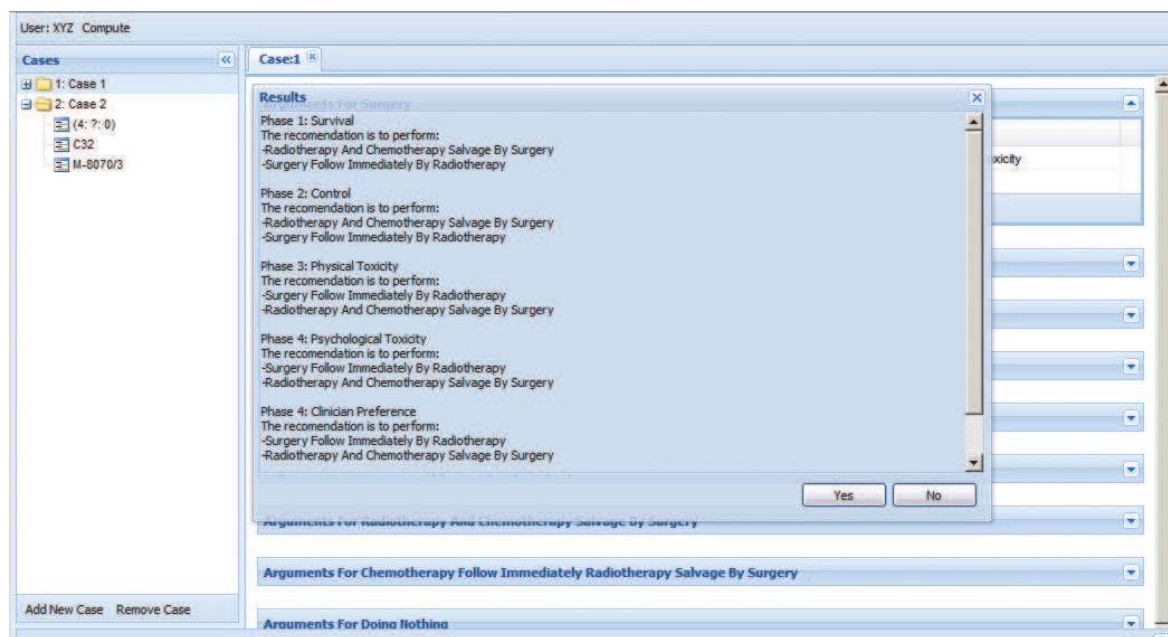
Figure 5.6: Recommendations

trial, several issues were identified. These issues falls into two categories: usability of the user interface and performance of the argumentation engine.

During an original execution of the tool, we found that when computing recommendations, the tool took sometime to return a decision (in some instances, several hours). This issue was address by limiting the scope to the available nine therapy choices rather than allowing the system to compute all therapy choices including non-existing ones. Although these non-existing therapy choices are valid with respect to the provided arguments, they are not valid with respect to the clinical scenario. This is directly attribute to the incompleteness of the background knowledge-base. However, if we were to assume that the background knowledge-base is complete, the discovery of new modalities suggests new treatment approaches. This is an interesting and somewhat surprising development that warrants future investigation.

Due to the technology chosen, no state information is retained. This restricted the flexibility of the application as the front-end application has to be left executing to wait for a response from the computationally expensive decision generation process. A more heavy weighted approach such as the use of Java programming language deploying the Java Platform, Enterprise Edition (Java EE) specification can be deployed to resolve this issue. However, this approach is against the original light weight web 2.0 philosophy. The move to Java will also potentially allow the use of existing argumentation engine such as Dungine [139]. However, modification to Dungine will need to be made as Dungine does not explicitly handle

preferences and the use of preference values in ranking extensions. By using Enterprise Java-Beans (EJB) and Web Services, users can periodically check-in with the service to check for updates.

The usability of the user interface can be addressed with a redesign given input from the user. Given the possible move to the use of Java and Web Services, more choices for the development of rich internet applications is now available. The use of technology such as HTML5, Adobe Flash or JavaFX becomes more feasible.

In following section, we will present a summary of the main concepts this discussed this chapter.

5.4 Summary

In this chapter, we have presented applications for the preferences-based accrual argumentation framework and the mixed-initiative argumentation framework. Within the section on applications of the preferences-based accrual argumentation framework, we firstly explored how the credibility of the source of an argument can be brought to bear in the argumentation machinery in two distinct settings. Firstly, we have introduced the semi-ring based accrual argumentation framework ($PAAF_{SR}$) and explored how credibility can be leveraged in an abstract argumentation framework (similar to those discussed in chapter 3). This framework is a modified version of the preferences-based accrual argumentation framework ($PAAF$). These abstract values are used to capture notions of source credibility and trust.

Secondly, we considered arguments with internal structure and presented the source-sensitive argumentation system ($SSAS$) as an instance of the the semi-ring based accrual argumentation framework. Utilising several examples, we have illustrated the ability for $PAAF_{SR}$ to perform source-sensitive argumentation and highlighted some of the unique features of the source-sensitive argumentation system ($SSAS$).

Furthermore, we presented a discussion on the applicability of this approach in areas such as defeasible argumentation and multi-agent negotiation.

Within section on applications of the mixed-initiative argumentation framework, we firstly presented an example of multi-disciplinary team session on cancer treatment. This example highlights the need for a clinical decision support tool that allows for the revision of the background knowledge-base over a sequence of decisions. We further highlighted the

complexity of the problem domain in support of the use of the frameworks proposed in this dissertation. The complexity of the domain is not only attributed to the different combination of treatment modalities, but also the different aspects of patient care criteria as well as the different sources of evidence required for supporting a treatment decision.

In the subsequent sections, we have described the construction, illustrated the use and evaluation of Just-Clinical. We have also discussed the issues encountered when using Just-Clinical. Of particular importance is the complexity of generation of all modalities, highlighting the difficulties in modality discovery. The proposed resolution is to restrict the tool to only consider modalities provided by the clinicians.

The next chapter will present the conclusions and future works.

Part III

Conclusion and Future Works

6

Conclusion

"This is not the end. This is not even the beginning of the end.
It is, instead, the end of the beginning."
– Winston Churchill

6.1   Conclusion

The thrust for this research is inspired by the fact that current available models of argumentation do not satisfactory capture and utilise preferences. Although this concept is certainly not new, it still provides a valuable insight despite the abundance presented argumentation models. This demonstrates that the study of argumentation is an interesting venture. Such a study provides methods and techniques for addressing real-world problems. The aim of this dissertation is to firstly, illustrate the importance of preferences in the accrual of argumentation and allowing this use of preferences to redefine the notion of acceptability criteria captured in existing argumentation. Secondly, to illustrate that an argumentation system

should not be static and introduce the notion of argumentation theory revision. The revision of an argumentation theory is based on the revision of the conflict and preference relation.

Firstly, we have introduced the abstract preference-based accrual argumentation framework ($PAAF$), illustrated the ability for the framework to perform accrual of arguments and highlighted some of the unique features of this framework. We have presented a discussion and comparison illustrating the differences between the abstract argumentation framework (AF) [52], the value-based argumentation framework (VAF) [22] and $PAAF$ as well as a discussion addressing issues such as argument source, credibility and trust as illustrated in the motivating example 1.2.3 and 1.2.4.

Secondly, we have presented a mixed-initiative argumentation framework. Within this framework, we have motivated the need for a mixed-initiative argumentation framework where argumentation theory is revised due to a sequence of decisions. By utilising mixed-initiative interaction, the argumentation theory can be revised to reflect the changes in knowledge in the "real-world". We have proposed a collection of procedures to perform such revision. Furthermore, we have proposed a set of properties governing the desirable behaviour for such argument theory change operator as well as shown that our procedures satisfies these properties. We also presented a discussion on the capabilities of such a framework focusing on the different combination of different structure in arguments and the presence/absence of a guiding knowledge-base.

Thirdly, we have presented a source-sensitive argumentation system. The source-sensitive argumentation system is viewed as an instance of the abstract preference-based accrual argumentation framework ($PAAF$). We have presented the vocabulary and the machinery that allow for computing the strength for sets of arguments. This vocabulary and machinery allows the system to address a range of problems. The aim of this system is to associate source with arguments and highlight the influence argumentation source have on the argumentation process. This associate of source to arguments allows for notions of credibility and trust to be captured in an argumentation system and for these notions to influence the outcome of the argumentation process

Finally, we presented a mixed-initiative argumentation system (Just-Clinical). We presented an example of multi-disciplinary team discussion within the medical domain, highlighting the need for a clinical decision support tool that allows for revision of the background knowledge-base over a sequence of decisions. The contribution of this dissertation is captured in the following key points:

1. Illustrated that argumentation provides methods and techniques for addressing problems that have no definitive "correct" answers or solutions.

2. Introduced a different utilisation of preference values in an abstract argumentation framework. Preferences are used for breaking ties between conflicting arguments and performing accrual of arguments.

3. Utilised preference values as informed representation of credibility and trust. Provided the vocabulary and machinery to address issues where information ownership or credibility plays an important role.

4. Highlighted the need to evolve an argumentation theory especially in situations where:

   (a) the argumentation system has been deployed in a dynamic environment.

   (b) the argumentation theory is incomplete or inaccurate (either incorrect or not-specific enough).

   (c) it is too difficult (time consuming, unable to obtain precision or knowledge) to elicit the knowledge required to perform reasoning.

5. Introduced the notion of decision-justification as an approach to evolve or correct the argumentation theory.

6. Proposed some desirable properties of the decision justification process.

7. Demonstrated the use of mixed-initiative argumentation in a clinical group decision support setting.

6.2 Future Works

In this section, we will present some future works and area of applications. There are three areas and directions of immediate interest: clinical decision support, distributed Constraint satisfaction and optimisation, requirements engineering.

6.2.1 Mixed-Initiative Argumentation for Modalities Discovery

As highlighted in chapter 5.3, the ability for the mixed-initiative argumentation system to discover new treatment modalities is an important issue. However, the issue lies with the

ability to collecting a sufficient amount of background knowledge such that the generated result is meaningful. One approach is for the mixed-initiative argumentation system to directly query pre-processed version of medical repository such as PubMed, hence allowing the system to draw upon a greater body of evidence. However, this requires further investigation as most evidence stored in such repositories are not design for reasoning by computational machines. One key point to highlight is that the mixed-initiative interaction machinery naturally lends itself to the accumulation of this background knowledge over a period of usage. Hence, given enough time, the background knowledge will eventually be more complete. However, further investigation is required.

6.2.2 Argumentation in Distributed Constraint Satisfaction and Dynamic Distributed Constraint Optimisation Problems

Recent work in distributed constraint satisfaction algorithms [68–72] is built upon the theoretical underpinnings described in this dissertation. Support-Based Distributed Search (SBDS) is a distributed constraint satisfaction algorithm in which agents communicate via arguments, maintaining a simple notion of credibility between agents.

The argument structure of SBDS is domain specific and permitting two categories of arguments. As it deviates from the argumentation structure representation in this dissertation (with distinction between facts and assumptions). Below is a loose description of the argument structure envisioned for SBDS.

Definition 6.2.1. *An SBDS-argument is a pair $\langle Prem, Con \rangle$, and belongs to one of the following two categories:*

1. *isgoods (variable-value assignment proposals)*

   - *Prem - an ordered sequence of variable-value assignments*
   - *Con - the variable-value assignment for the agent stating the argument*

2. *nogoods (variable-value assignment rejections)*

   - *Prem - a set of variable-value assignments which are not permitted*
   - *Con - exact copy of the premise*

As the argument structure of SBDS differs from that discussion in this dissertation, a domain-specific conflict and attacks relation is needed. The spirit of these relation remains the same

as those provided in this dissertation. This further strengthen the claim that our proposed system complements and provides assistance to solving problems in a whole range of different domains.

Definition 6.2.2. *Given* β, γ *are SBDS arguments, then* $conflict(\beta, \gamma)$ *if and only if the simultaneous application of* Con_β *and* Con_γ *result in domain wipeout for some variable.*

Definition 6.2.3. *Given* β, γ *are SBDS arguments, then* $attack(\beta, \gamma)$ *if and only if* $conflict(\beta, \gamma)$ *and either* β *is a nogood, or both are isgoods and the* $Prem_\beta \cup Con_\beta$ *is a 'stronger' assignment than* $Prem_\gamma \cup Con_\gamma$.

The term 'stronger' is specific to SBDS and describes the use of a monotonic function for comparing sequences of variable-value assignments. By defining attack in this way we demonstrate that the source-sensitive argumentation system is not limited to simple n-valued logics, but can be applied more generically. Note that by definition 6.2.3, a 'nogood' argument will never come under attack. This is appropriate as a 'nogood' is always a deductive argument, whereas an 'isgood' is an inductive argument of varying degrees of confidence. The concept of credibility is used very simply in SBDS, but in even limited capacity causes the same agent-behaviour as we predicted. There are only 3 levels of credibility in SBDS.

As this is a loose description of the argument structure envisioned for SBDS, further exploration is required.

6.2.3 Mixed-Initiative Argumentation and Requirements Engineering

Recall from chapter 3 that our ultimate goal is to create a framework that capture arguments and their associated preferences, this results in the ability to associate credibility as well as ownership to arguments. The proposal here is to implement the argumentation system within a fine grained multi-agent executable environment which co-exists with the *i** model framework [153–156]. A very abstract summary of the envisioned benefits of such a co-existence given in the following points:

- Early-phase requirements are specified in *i** models.

- These are progressively refined to late-phase (detailed) requirements represented via a combination of *i** models with Formal-Tropos annotations (*creation condition, fulfillment condition, etc.*) and AgentSpeak(L-SG) agents. We visualise the two components

of such requirements specifications co-evolving. Such co-evolution would entail independent development of both components of these specifications, while maintaining some loose consistency links between them.

- Further refinements of these specifications lead to architectural models, in a manner paralleling the TROPOS methodology.

- We view such specifications as stake-holder *goal-intention-context models*. We visualise maintaining such models throughout the software life-cycle. Such models would be used to perform verification, validation and trade-off analysis on change requests, in particular, requirements change requests (which have been recognised in the literature as the most expensive and difficult to deal with).

Utilising this approach, we are able to explore two alternative approaches to stakeholder negotiation:

1. Negotiation via argumentation. Automated negotiation is a novel approach for resolving conflicts. The minimal structure of a requirement lends itself to argumentation where arguments are a pair consisting of a premise and a conclusion. In the case of requirements, the rationale is the premise and the conclusion is the requirement. Arguments can be generated from the list of requirements.

2. Negotiation via belief merging. This approach is based on semantic accounts of how preference specifications of a society of agents can be aggregated into a combined preference relation.

However, further investigation is required.

Bibliography

[1] Mitchell Ai-Chang, John L. Bresina, Leonard Charest, Adam Chase, Jennifer Cheng jung Hsu, Ari K. Jónsson, Bob Kanefsky, Paul H. Morris, Kanna Rajan, Jeffrey Yglesias, Brian G. Chafin, William C. Dias, and Pierre F. Maldague. Mapgen: Mixed-initiative planning and scheduling for the mars exploration rover mission. *IEEE Intelligent Systems*, 19(1):8–12, 2004.

[2] Carlos E. Alchourron, Peter Gärdenfors, and David Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982.

[3] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.*, 50(2):510–530, 1985.

[4] James F. Allen. Mixed-initiative interaction. *IEEE Intelligent Systems*, 14(5):14–23, 1999.

[5] Simon Ambler. A categorial approach to the semantics of argumentation. *Mathematical Structures in Computer Science*, 6(2):167–188, 1996.

[6] Leila Amgoud. Using preferences to select acceptable arguments. In *ECAI*, pages 43–44, 1998.

[7] Leila Amgoud and Claudette Cayrol. Integrating preference orderings into argument-based reasoning. In Dov M. Gabbay, Rudolf Kruse, Andreas Nonnengart, and Hans Jürgen Ohlbach, editors, *ECSQARU-FAPR*, volume 1244 of *Lecture Notes in Computer Science*, pages 159–170. Springer, 1997.

[8] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In *UAI*, pages 1–7, 1998.

[9] Leila Amgoud and Claudette Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *J. Autom. Reasoning*, 29(2):125–169, 2002.

[10] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Ann. Math. Artif. Intell.*, 34(1-3):197–215, 2002.

[11] Leila Amgoud, Claudette Cayrol, and Daniel Le Berre. Comparing arguments using preference ordering for argument-based reasoning. In *ICTAI*, pages 400–403, 1996.

[12] Leila Amgoud, Yannis Dimopoulos, and Pavlos Moraitis. Making decisions through preference-based argumentation. In Brewka and Lang [29], pages 113–123.

[13] Leila Amgoud and Simon Parsons. Agent dialogues with conflicting preferences. In *ATAL*, pages 190–205, 2001.

[14] Leila Amgoud and Henri Prade. Explaining qualitative decision under uncertainty by argumentation. In *AAAI*. AAAI Press, 2006.

[15] Leila Amgoud and Henri Prade. Using arguments for making and explaining decisions. *Artif. Intell.*, 173(3-4):413–436, 2009.

[16] Leila Amgoud and Srdjan Vesic. On revising argumentation-based decision systems. In Claudio Sossai and Gaetano Chemello, editors, *ECSQARU*, volume 5590 of *Lecture Notes in Computer Science*, pages 71–82. Springer, 2009.

[17] Leila Amgoud and Srdjan Vesic. A formal analysis of the outcomes of argumentation-based negotiations. In Liz Sonenberg, Peter Stone, Kagan Tumer, and Pinar Yolum, editors, *AAMAS*, pages 1237–1238. IFAAMAS, 2011.

[18] Leila Amgoud and Srdjan Vesic. Two roles of preferences in argumentation frameworks. In Weiru Liu, editor, *ECSQARU*, volume 6717 of *Lecture Notes in Computer Science*, pages 86–97. Springer, 2011.

[19] Grigoris Antoniou. A tutorial on default reasoning. *Knowl. Eng. Rev.*, 13(3):225–246, 1998.

[20] Kenneth J. Arrow. *Social Choice and Individual Values, Second edition (Cowles Foundation Monographs Series)*. Yale University Press, 2nd edition, September 1970.

[21] Katie Atkinson, Trevor J. M. Bench-Capon, and Sanjay Modgil. Argumentation for decision support. In Stéphane Bressan, Josef Küng, and Roland Wagner, editors, *DEXA*, volume 4080 of *Lecture Notes in Computer Science*, pages 822–831. Springer, 2006.

[22] Trevor J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[23] P. Besnard and A. Hunter. Argumentation Based on Classical Logic. In Simari, G. & Rahwan, I., editor, *Argumentation in Artificial Intelligence*, page 133. Springer US, 2009.

[24] Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Journal of Artificial Intelligence*, 93:63–101, 1997.

[25] Jean-Rémi Bourguet, Leila Amgoud, and Rallou Thomopoulos. Towards a unified model of preference-based argumentation. In Sebastian Link and Henri Prade, editors, *FoIKS*, volume 5956 of *Lecture Notes in Computer Science*, pages 326–344. Springer, 2010.

[26] John L. Bresina, Ari K. Jónsson, Paul H. Morris, and Kanna Rajan. Mixed-initiative activity planning for mars rovers. In Leslie Pack Kaelbling and Alessandro Saffiotti, editors, *IJCAI*, pages 1709–1710. Professional Book Center, 2005.

[27] John L. Bresina and Paul H. Morris. Mixed-initiative planning in space mission operations. *AI Magazine*, 28(2):75–88, 2007.

[28] Gerhard Brewka. Dynamic argument systems: A formal model of argumentation processes based on situation calculus. *Journal of Logic and Computation*, 11(2):257–282, 2001.

[29] Gerhard Brewka and Jérôme Lang, editors. *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*. AAAI Press, 2008.

[30] Mark H. Burstein, George Ferguson, and James F. Allen. Integrating agent-based mixed-initiative control with an existing multi-agent planning system. In *ICMAS*, pages 389–390. IEEE Computer Society, 2000.

[31] Mark H. Burstein and Drew V. McDermott. Chapter 17 issues in the development of human-computer mixed-initiative planning. In Barbara Gorayska and Jacob L. Mey, editors, *Cognitive Technology In Search of a Humane Interface*, volume 113 of *Advances in Psychology*, pages 285 – 303. North-Holland, 1996.

[32] Allan Caine and Robin Cohen. Mits: A mixed-initiative intelligent tutoring system for sudoku. In Luc Lamontagne and Mario Marchand, editors, *Canadian Conference*

on AI, volume 4013 of *Lecture Notes in Computer Science*, pages 550–561. Springer, 2006.

[33] Daniela V. Carbogim. *Dynamics in Formal Argumentation*. PhD thesis, University of Edinburgh. College of Science and Engineering. School of Informatics, 2000.

[34] Daniela V. Carbogim and Renata Wassermann. Full acceptance via argumentation. In Maria Carolina Monard and Jaime Simão Sichman, editors, *IBERAMIA-SBIA 2000 Open Discussion Track*, pages 127–136. ICMC/USP, 2000.

[35] Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Revision of an argumentation system. In Brewka and Lang [29], pages 124–134.

[36] Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *J. Artif. Intell. Res. (JAIR)*, 38:49–84, 2010.

[37] Chee Fon Chang, Aditya Ghose, and Andrew Miller. Mixed-initiative argumentation: A framework for justification management in clinical group decision support. In *The Uses of Computational Argumentation, AAAI Fall Symposium*, pages 14 –19, November 2009.

[38] Chee Fon Chang, Peter Harvey, and Aditya Ghose. Source sensitive argumentation system. In Yannis Manolopoulos, Joaquim Filipe, Panos Constantopoulos, and José Cordeiro, editors, *ICEIS*, pages 39–46, 2006.

[39] Chee Fon Chang, Peter Harvey, and Aditya K. Ghose. Combining credibility in a source sensitive argumentation system. In *SETN*, pages 478–481, 2006.

[40] Chee Fon Chang, Andrew Miller, and Aditya Ghose. Mixed-initiative argumentation: Group decision support in medicine. In *Second International ICST Conference on Electronic Healthcare for the 21st century*, pages 43–50, October 2009.

[41] Carlos Iván Chesñevar, Ana Gabriela Maguitman, and Ronald Prescott Loui. Logical models of argument. *ACM Comput. Surv.*, 32(4):337–383, 2000.

[42] Jennifer Chu-Carroll and Michael K. Brown. An evidential model for tracking initiative in collaborative dialogue interactions. *User Model. User-Adapt. Interact.*, 8(3-4):215–254, 1998.

[43] Robin Cohen, Coralee Allaby, Christian Cumbaa, Mark Fitzgerald, Kinson Ho, Bowen Hui, Celine Latulipe, Fletcher Lu, Nancy Moussa, David Pooley, Alex Qian, and Saheem Siddiqi. What is initiative? *User Model. User-Adapt. Interact.*, 8(3-4):171–214, 1998.

[44] Robin Cohen and Michael Fleming. Adjusting the autonomy in mixed-initiative systems by reasoning about interaction. In Henry Hexmoor, Cristiano Castelfranchi, Rino Falcone, and Gerhard Weiss, editors, *Agent Autonomy*, volume 7 of *Multiagent Systems, Artificial Societies, and Simulated Organizations*, pages 137–158. Springer US, 2003.

[45] Michael T. Cox and Manuela M. Veloso. Controlling for unexpected goals when planning in a mixed-initiative setting. In Ernesto Costa and Amílcar Cardoso, editors, *EPIA*, volume 1323 of *Lecture Notes in Computer Science*, pages 309–318. Springer, 1997.

[46] Michael T. Cox and Chen Zhang. Mixed-initiative goal manipulation. *AI Magazine*, 28(2):62–74, 2007.

[47] Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artif. Intell.*, 89(1-2):1–29, 1997.

[48] Prasun Dewan, Jonathan Grudin, and Eric Horvitz. Towards mixed-initiative access control. In *CollaborateCom*, pages 64–71. IEEE, 2007.

[49] Sylvie Doutre, Trevor J. M. Bench-Capon, and Paul E. Dunne. Determining preferences through argumentation. In *AI*IA*, pages 98–109, 2005.

[50] Sylvie Doutre, Trevor J. M. Bench-Capon, and Paul E. Dunne. Explaining preferences with argument positions. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005*, pages 1560–1561, 2005.

[51] Sylvie Doutre and Jérôme Mengin. Preferred extensions of argumentation frameworks: Query answering and computation. In Rajeev Goré, Alexander Leitsch, and Tobias Nipkow, editors, *IJCAR*, volume 2083 of *Lecture Notes in Computer Science*, pages 272–288. Springer, 2001.

[52] Phan Minh Dung. An argumentation-theoretic foundations for logic programming. *Journal of Logic Programming*, 22(2):151–171, 1995.

[53] Phan Minh Dung and Tran Cao Son. An argument-based approach to reasoning with specificity. *Journal of Artificial Intelligence*, 133(1-2):35–85, 2001.

[54] George Ferguson and James F. Allen. Arguing about plans: Plan representation and reasoning for mixed-initiative planning. In *AIPS*, pages 43–48, 1994.

[55] George Ferguson and James F. Allen. Trips: The rochester interactive planning system. In *AAAI/IAAI*, pages 906–907, 1999.

[56] George Ferguson and James F. Allen. Mixed-initiative systems for collaborative problem solving. *AI Magazine*, 28(2):23–32, 2007.

[57] George Ferguson, James F. Allen, and Bradford W. Miller. Trains-95: Towards a mixed-initiative planning assistant. In *AIPS*, pages 70–77, 1996.

[58] Eduardo L. Fermé. On the logic of theory change: Contraction without recovery. *Journal of Logic, Language and Information*, 7(2):127–137, 1998.

[59] Alberto Finzi and Andrea Orlandini. Human-robot interaction through mixed-initiative planning for rescue and search rovers. In *AI*IA*, pages 483–494, 2005.

[60] John Fox, David Glasspool, Dan Grecu, Sanjay Modgil, Matthew South, and Vivek Patkar. Argumentation-based inference and decision making–a medical perspective. *IEEE Intelligent Systems*, 22(6):34–41, 2007.

[61] Peter Gärdenfors. Belief revision and knowledge representation. In Yoav Shoham, editor, *TARK*, page 117. Morgan Kaufmann, 1996.

[62] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States (Studies in Logic: Mathematical Logic and Foundations)*. College Publications, jun 2008.

[63] Peter Gärdenfors. Notes on the history of ideas behind agm. *Journal of Philosophical Logic*, 40:115–120, 2011.

[64] Michael A. Gilbert. The delimitation of argument. *Inquiry*, 15(1):63–75, 1995.

[65] Jaap Hage. A theory of legal reasoning and a logic to match. *Artif. Intell. Law*, 4(3-4):199–273, 1996.

[66] Susan Haller, Susan McRoy, and Alfred Kobsa. *Computational Models of Mixed-Initiative Interaction*. Springer Publishing Company, Incorporated, 1st edition, 2007.

[67] William L. Harper. Rational conceptual change. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1976:462–494, 1976.

[68] Peter A. Harvey, Chee Fon Chang, and Aditya K. Ghose. Practical application of support-based distributed search. In *ICTAI*, pages 34–38, 2005.

[69] Peter A. Harvey, Chee Fon Chang, and Aditya K. Ghose. Support-based distributed search. In Amnon Meisels, editor, *Proceedings of the 6th International Workshop on Distributed Constraint Reasoning*, July - August 2005. http://www.dsl.uow.edu.au/people/harvey/dcr05.pdf.

[70] Peter A. Harvey, Chee Fon Chang, and Aditya K. Ghose. Simple support-based distributed search. In *Canadian Conference on AI*, pages 159–170, 2006.

[71] Peter A. Harvey, Chee Fon Chang, and Aditya K. Ghose. Support-based distributed search: A new approach for multiagent constraint processing. In Nicolas Maudet, Simon Parsons, and Iyad Rahwan, editors, *ArgMAS*, volume 4766 of *Lecture Notes in Computer Science*, pages 91–106. Springer, 2006.

[72] Peter A. Harvey, Chee Fon Chang, and Aditya K. Ghose. Support-based distributed search: a new approach for multiagent constraint processing. In *Proceedings of the 5th International Joint Conference on Autonomous Agents & Multi Agent Systems*, pages 377–383, 2006.

[73] Eric Horvitz. Principles of mixed-initiative user interfaces. In *CHI*, pages 159–166, 1999.

[74] Eric Horvitz. Reflections on challenges and promises of mixed-initiative interaction. *AI Magazine*, 28(2):13–22, 2007.

[75] Nicholas R. Jennings, Peyman Faratin, Alessio R. Lomuscio, Simon Parsons, Michael Wooldridge, and Carles Sierra. Automated negotiation: prospects, methods and challenges. *Intern. J. of Group Decision and Negotiation*, 10(2):199–215, 2001.

[76] Nicholas R. Jennings, Simon Parsons, Pablo Noriega, and Carles Sierra. On argumentation-based negotiation. In *Proceedings of the International Workshop on Multi-Agent Systems*, Boston, USA, 1998.

[77] Souhila Kaci and Leendert van der Torre. Preference-based argumentation: Arguments supporting multiple values. *Int. J. Approx. Reasoning*, 48(3):730–751, 2008.

[78] Souhila Kaci, Leendert W. N. van der Torre, and Emil Weydert. Acyclic argumentation: Attack = conflict + preference. In Gerhard Brewka, Silvia Coradeschi, Anna Perini, and Paolo Traverso, editors, *ECAI*, pages 725–726. IOS Press, 2006.

[79] Antonis C. Kakas, Robert A. Kowalski, and Francesca Toni. Abductive logic programming. *J. Log. Comput.*, 2(6):719–770, 1992.

[80] Antonis C. Kakas and Paolo Mancarella. Abduction and abductive logic programming. In *ICLP*, pages 18–19, 1994.

[81] Nikos Karacapilidis and Dimitris Papadias. Computer supported argumentation and collaborative decision making: the hermes system. *Inf. Syst.*, 26(4):259–277, 2001.

[82] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In *KR*, pages 387–394, 1991.

[83] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. *Artif. Intell.*, 52(3):263–294, 1992.

[84] Arthur E. Kirkpatrick, Bistra Dilkina, and William S. Havens. A framework for designing and evaluating mixed-initiative optimization systems. In *In Proceedings of the ICAPS-05 Workshop on Mixed-initiative Planning and Scheduling*, 2005.

[85] Sébastien Konieczny and Ramón Pino Pérez. On the logic of merging. In *KR*, pages 488–498, 1998.

[86] Sarit Kraus, Katia P. Sycara, and Amir Evenchik. Reaching agreements through argumentation: A logical model and implementation. *Journal of Artificial Intelligence*, 104(1-2):1–69, 1998.

[87] Paul Krause, Simon Ambler, Morten Elvang-Gøransson, and John Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11:113–131, 1995.

[88] Isaac Levi. *The Fixation of Belief and its Undoing: Changing Beliefs through Inquiry*. Cambridge University Press, oct 1991.

[89] Fangzhen Lin and Yoav Shoham. Argument systems: A uniform basis for nonmonotonic reasoning. In *KR*, pages 245–255, 1989.

[90] Ronald P. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 3:100–106, 1987.

[91] Ronald Prescott Loui, Jeff Norman, Jon Olson, and Andrew Merrill. A design for reasoning with policies, precedents, and rationales. In *ICAIL*, pages 202–211, 1993.

[92] Mauro Javier Gómez Lucero, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. On the accrual of arguments in defeasible logic programming. In *IJCAI*, pages 804–809, 2009.

[93] David Makinson. On the status of the postulate of recovery in the logic of theory change. *Journal of Philosophical Logic*, 16:383–394, 1987.

[94] John McCarthy. Addendum: Circumscription and other non-monotonic formalisms. *Artif. Intell.*, 13(1-2):171–172, 1980.

[95] John McCarthy. Circumscription - a form of non-monotonic reasoning. *Artif. Intell.*, 13(1-2):27–39, 1980.

[96] Drew V. McDermott. Nonmonotonic logic ii: Nonmonotonic modal theories. *J. ACM*, 29(1):33–57, 1982.

[97] Drew V. McDermott and Jon Doyle. Non-monotonic logic i. *Artif. Intell.*, 13(1-2):41–72, 1980.

[98] Thomas Meyer. Merging epistemic states. In *PRICAI*, pages 286–296, 2000.

[99] Thomas Meyer. On the semantics of merging. *CoRR*, cs.AI/0003015, 2000.

[100] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, 173(9-10):901–934, 2009.

[101] Robert C. Moore. Semantical considerations on nonmonotonic logic. *Artif. Intell.*, 25(1):75–94, 1985.

[102] Alice M. Mulvehill, Brett Benyo, Michael T. Cox, and Renu Kurien Bostwick. Expectation failure as a basis for agent-based model diagnosis and mixed initiative model adaptation during anomalous plan execution. In Veloso [143], pages 489–494.

[103] Shmuel Nitzan. *Collective Preference and Choice*. Cambridge University Press, 1 edition, January 2010.

[104] Donald Nute. Defeasible reasoning and decision support systems. *Decision Support Systems*, 4:97–110, 1988.

[105] Donald Nute. A nonmonotonic logic based on conditional logic. Technical report acmc 01-0007, Advanced Computational Methods Center, University of Georgia, Athens, GA, 1996.

[106] Simon Parsons, Peter McBurney, Elizabeth Sklar, and Michael Wooldridge. On the relevance of utterances in formal inter-agent dialogues. In Edmund H. Durfee, Makoto Yokoo, Michael N. Huhns, and Onn Shehory, editors, *AAMAS*, page 240. IFAAMAS, 2007.

[107] Simon Parsons, Carles Sierra, and Nicholas R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[108] Chaim Perelman. *Justice, law, and argument : essays on moral and legal reasoning / Ch. Perelman ; with an introd. by Harold J. Berman.* D. Reidel Pub. Co. ; sold and distributed in the U.S.A. and Canada by Kluwer Boston, Dordrecht, Holland ; Boston : Hingham, MA :, 1980.

[109] John L. Pollock. The structure of epistemic justification. *American Philosophical Quarterly*, monograph series 4:62–78, 1970.

[110] John L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.

[111] John L. Pollock. Interest driven suppositional reasoning. *Journal on Automated Reasoning*, 6:419–461, 1990.

[112] John L. Pollock. Self-defeating arguments. *Minds and Machines*, 1(4):367–392, 1991.

[113] John L. Pollock. A theory of defeasible reasoning. *International Journal on Intelligent Systems*, 6:33–54, 1991.

[114] John L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

[115] John L. Pollock. Justification and defeat. *Artificial Intelligence*, 67:337–407, 1994.

[116] John L. Pollock. *Cognitive Carpentry: A Blueprint for how to Build a Person.* MIT Press, Cambridge, MA, USA, 1995.

[117] John L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133(1-2):233–282, 2001.

[118] D.L. Poole, R. Goebel, and R. Aleliunas. *Theorist: a logical reasoning system for defaults and diagnosis.* Research report (University of Waterloo. Faculty of Mathematics). Faculty of Mathematics, University of Waterloo, 1986.

[119] Henry Prakken. An argumentation framework in default logic. *Annals of Mathematics and Artificial Intelligence*, 9(1-2):93–132, 1993.

[120] Henry Prakken. From logic to dialectics in legal argument. In *ICAIL*, pages 165–174, 1995.

[121] Henry Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127(1):187–219, 2001.

[122] Henry Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *ICAIL*, pages 85–94. ACM, 2005.

[123] Henry Prakken and Giovanni Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. *Artif. Intell. Law*, 4(3-4):331–368, 1996.

[124] Henry Prakken and Giovanni Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artif. Intell. Law*, 6(2-4):231–287, 1998.

[125] Henry Prakken and Gerard Vreeswijk. *Logics for Defeasible Argumentation*, volume 4 of *Handbook of Philosophical Logic*, pages 218–319. Kluwer Academic Publishers, second edition edition, 2002.

[126] Iyad Rahwan, Pavlos Moraitis, and Chris Reed, editors. *Argumentation in Multi-Agent Systems, First International Workshop, ArgMAS 2004, New York, NY, USA, July 19, 2004, Revised Selected and Invited Papers*, volume 3366 of *Lecture Notes in Computer Science*. Springer, 2005.

[127] Iyad Rahwan, Liz Sonenberg, and Peter McBurney. Bargaining and argument-based negotiation: Some preliminary comparisons. In Rahwan et al. [126], pages 176–191.

[128] William Rehg, Peter McBurney, and Simon Parsons. Computer decision-support systems for public argumentation: assessing deliberative legitimacy. *AI Soc.*, 19(3):203–228, 2005.

[129] Raymond Reiter. A logic for default reasoning. *Artif. Intell.*, 13(1-2):81–132, 1980.

[130] Charles Rich and Candace L. Sidner. Collagen: A collaboration manager for software interface agents. *User Model. User-Adapt. Interact.*, 8(3-4):315–350, 1998.

[131] Charles Rich, Candace L. Sidner, and Neal Lesh. Collagen: Applying collaborative discourse theory to human-computer interaction. *AI Magazine*, 22(4):15–26, 2001.

[132] Hans Rott. A nonmonotonic conditional logic for belief revision. part 1: Semantics and logic of simple conditionals. In André Fuhrmann and Michael Morreau, editors, *The Logic of Theory Change*, volume 465 of *Lecture Notes in Computer Science*, pages 135–181. Springer, 1989.

[133] Hans Rott. Preferential belief change using generalized epistemic entrenchment. *Journal of Logic, Language and Information*, 1(1):45–78, 1992.

[134] Hans Rott. Coherence and conservatism in the dynamics of belief ii: Iterated belief change without dispositional coherence. *J. Log. Comput.*, 13(1):111–145, 2003.

[135] Sonia V. Rueda, Alejandro J. Garcia, and Guillermo R. Simari. Argument-based negotiation among BDI agents. *Journal of Computer Science and Technology*, 2(7), oct 2002.

[136] Carles Sierra, Nicholas R. Jennings, Pablo Noriega, and Simon Parsons. A framework for argumentation-based negotiation. In Munindar P. Singh, Anand S. Rao, and Michael Wooldridge, editors, *Proceedings of the 4th International Workshop on Agent Theories, Architectures, and Languages*, pages 177–192, 1997.

[137] Teresa Sikorski and James F. Allen. A task-based evaluation of the trains-95 dialogue system. In Elisabeth Maier, Marion Mast, and Susann LuperFoy, editors, *ECAI Workshop on Dialogue Processing in Spoken Language Systems*, volume 1236 of *Lecture Notes in Computer Science*, pages 207–220. Springer, 1996.

[138] Guillermo Ricardo Simari and Ronald Prescott Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artif. Intell.*, 53(2-3):125–157, 1992.

[139] Matthew South, Gerard Vreeswijk, and John Fox. Dungine: A java dung reasoner. In Philippe Besnard, Sylvie Doutre, and Anthony Hunter, editors, *COMMA*, volume 172 of *Frontiers in Artificial Intelligence and Applications*, pages 360–368. IOS Press, 2008.

[140] K. Suzumura. *Rational choice, collective decisions, and social welfare*. Cambridge University Press, 1983.

[141] Gheorghe Tecuci, Mihai Boicu, and Michael T. Cox. Seven aspects of mixed-initiative reasoning: An introduction to this special issue on mixed-initiative assistants. *AI Magazine*, 28(2):11–12, 2007.

[142] Frans H. van Eemeren, Rob Grootendorst, and Francisca Snoeck Henkemans. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Lawrence Erlbaum Associates, 1996.

[143] Manuela M. Veloso, editor. *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, 2007.

[144] Manuela M. Veloso, Alice M. Mulvehill, and Michael T. Cox. Rationale-supported mixed-initiative case-based planning. In *AAAI/IAAI*, pages 1072–1077, 1997.

[145] Bart H. Verheij. Reason based logic and legal knowledge representation. In I. Carr and A. Narayanan, editors, *Proceedings of the 4th National Conference on Law, Computers and Artificial Intelligence*, pages 154–165. University of Exeter, 1994.

[146] Bart H. Verheij. Accrual of arguments in defeasible argumentation. In *The proceedings of the 2nd Dutch/German Workshop on Nonmonotonic Reasoning, Delft University of Technology, Utrecht University, Also published as report SKBS/B3.A/95-01*, pages 217–224, 1995.

[147] Bart H. Verheij. *Rules, reasons, arguments: formal studies of argumentation and defeat.* Doctoral dissertation, University of Maastricht, 1996.

[148] Bart H. Verheij. A labeling approach to the computation of credulous acceptance in argumentation. In Veloso [143], pages 623–628.

[149] Bart H. Verheij, Jaap Hage, and H. Jaap van den Herik. An integrated view on rules and principles. *Artif. Intell. Law*, 6(1):3–26, 1998.

[150] Gerard A. W. Vreeswijk. Abstract argumentation systems. *Journal of Artificial Intelligence*, 90(1-2):225–279, 1997.

[151] Gerard A. W. Vreeswijk. Argumentation in bayesian belief networks. In Rahwan et al. [126], pages 111–129.

[152] Marilyn A. Walker and Steve Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. In *ACL*, pages 70–78, 1990.

[153] Eric S. K. Yu. *Modelling Strategic Relationships for Process Reengineering*. PhD thesis, Department of Computer Science, University of Toronto, 1995.

[154] Eric S. K. Yu and John Mylopoulos. From e-r to "a-r" - modelling strategic actor relationships for business process reengineering. In Pericles Loucopoulos, editor, *ER*, volume 881 of *Lecture Notes in Computer Science*, pages 548–565. Springer, 1994.

[155] Eric S. K. Yu and John Mylopoulos. Understanding "why" in software process modelling, analysis, and design. In Bruno Fadini, Leon J. Osterweil, and Axel van Lamsweerde, editors, *ICSE*, pages 159–168. IEEE Computer Society / ACM Press, 1994.

[156] Eric S. K. Yu and John Mylopoulos. From e-r to "a-r" - modelling strategic actor relationships for business process reengineering. *Int. J. Cooperative Inf. Syst.*, 4(2-3):125–144, 1995.

A

Procedures

The following are the formalisation of the procedures described in chapter 4. A total of four procedures are presented in detail. The four procedures are as listed:

- General decision justification procedure.

- Preference-modifying decision justification procedure.

- Conflict-modifying decision justification procedure.

- Conflict-modifying and preference-modifying decision justification procedure.

Procedure A.1: General Decision Justification Procedure.

Inputs: $AT = \langle AR, Conf, Pref \rangle$, and a user-defined extension S.
Outputs: $AT' = \langle AR', Conf', Pref' \rangle$
Begin
  $Ext_{AT} \leftarrow$ ***Generate***(AT)
  While $(Ext_{AT} \setminus \{S\}) \neq \emptyset$ **Do**
    If AR *is S-complete* **Then**
      If $S \in Ext_{AT}$ **Then**
        $Pref \leftarrow$ ***Pref-Mod***(AT, Ext_{AT}, S) ▷ call sub-procedure
      Else
        $Conf \leftarrow$ ***Conf-Mod***(AT, Ext_{AT}, S) ▷ call sub-procedure
      EndIf
    Else
      $AR \leftarrow AR \cup S$
    EndIf
    $Ext_{AT} \leftarrow$ ***Generate***(AT)
  EndWhile
  Return AT
End

Procedure A.2: Preference-modifying Decision Justification Procedure (*Pref-Mod*).

Inputs: $AT = \langle AR, Conf, Pref \rangle$, Ext_{AT} generated from AT, a user-defined extension S.
Outputs: A modified preference theory $Pref'$.
Begin
  $Pref' \leftarrow Pref$
  Forall $A_i \in (Ext_{AT} \setminus \{S\})$ **Do**
    If $pref(S, A_i) \notin Pref'$ **Then**
      $Pref' \leftarrow (Pref' \setminus \{pref(A_i, S)\}) \cup \{pref(S, A_i)\}$ ▷ Reverse preference
    EndIf
  EndForall
  Return $Pref'$
End

Procedure A.3: Conflict-modifying Decision Justification Procedure (*Conf-Mod*).

Inputs: $AT = \langle AR, Conf, Pref \rangle$, Ext_{AT} generated from AT, a user-defined extension S.
Outputs: A modified conflict theory $Conf'$.
Selection Functions: $f_a : 2^{AR} \longrightarrow AR$.
Begin

$Conf' \leftarrow Conf$
Forall $\alpha_i, \alpha_j \in S$ **Do**   ▷ Check S for self defeat
  If $deft(\alpha_i, \alpha_j) \in Deft$ **Then**
    $Conf' \leftarrow (Conf' \setminus (\{conf(\alpha_j, \alpha_i)\} \cup \{conf(\alpha_i, \alpha_j)\}))$
  EndIf
EndForall
$AR' \leftarrow (AR \setminus S)$ ▷ Remove S from consideration
Forall $deft(\alpha_i, \alpha_j) \in Deft$ **Do**
  If $\alpha_j \in AR'$ *and* $\alpha_i \in S$ **Then**   ▷ Identify defeated arguments
    $AR' \leftarrow AR' \setminus \{\alpha_j\}$ ▷ Ignore defeated arguments
    $S' \leftarrow S' \cup \{\alpha_i\}$ ▷ Construct a set of undefeated from S
  EndIf
EndForall
Forall $deft(\alpha_i, \alpha_j) \in Deft$ **Do**
  If $\alpha_i \in AR'$ *and* $\alpha_j \in S$ **Then**
    $\alpha_k = f_a(S' \cup \{\alpha_j\})$ ▷ non-deterministic selection
    $Conf' \leftarrow Conf' \cup \{conf(\alpha_k, \alpha_i)\}$
    If $\alpha_k = \alpha_j$ *and* $\neg(pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref$ *and* $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref)$ **Then**
      $Conf' \leftarrow Conf' \setminus \{conf(\alpha_i, \alpha_j)\}$
    EndIf
    $AR' \leftarrow AR' \setminus \{\alpha_i\}$
  EndIf
EndForall
Forall $\alpha_i \in AR'$ **Do**   ▷ Process non-conflicting arguments
  $\alpha_k = f_a(S)$ ▷ non-deterministic selection
  $Conf' \leftarrow (Conf' \setminus \{conf(\alpha_i, \alpha_k)\}) \cup \{conf(\alpha_k, \alpha_i)\}$
EndForall
Return $Conf'$

End
Non-deterministic selection function f_a describes a class of operator and can be replaced with a
function that select the $i^t h$ enumeration of a list of arguments or choices provided from user interaction.

Procedure A.4: Conflict and Preference Modifying Decision Justification Procedure

---

Inputs: $AT = \langle AR, Conf, Pref \rangle$, a user-defined extension S.
Outputs: A modified conflict-preference theory pair $\langle Conf', Pref' \rangle$
Selection Functions: $f_a : 2^{AR} \longrightarrow AR$.
Begin

  $Conf' \leftarrow Conf$
  $Pref' \leftarrow Pref$
  Forall $\alpha_i, \alpha_j \in S$ **Do**   ▷ Check S for self defeat
    If $deft(\alpha_i, \alpha_j) \in Deft$ **Then**
      If $\{conf(\alpha_j, \alpha_i), conf(\alpha_i, \alpha_j)\} \subseteq Conf'$ **Then**
        Select   ▷ non-deterministic selection
          $Conf' \leftarrow (Conf' \setminus (\{conf(\alpha_j, \alpha_i)\} \cup \{conf(\alpha_i, \alpha_j)\}))$
          or
          $Pref' \leftarrow Pref' \cup \{pref(\{\alpha_i\}, \{\alpha_j\})\} \cup \{pref(\{\alpha_j\}, \{\alpha_i\})\}$
        EndSelect
      Else
        $Conf' \leftarrow (Conf' \setminus (\{conf(\alpha_j, \alpha_i)\} \cup \{conf(\alpha_i, \alpha_j)\}))$
      EndIf
    EndIf
  EndForall
  $AR' \leftarrow (AR \setminus S)$ ▷ Remove S from consideration
  Forall $deft(\alpha_i, \alpha_j) \in Deft$ **Do**
    If $\alpha_j \in AR'$ *and* $\alpha_i \in S$ **Then**   ▷ Identify defeated arguments
      $AR' \leftarrow AR' \setminus \{\alpha_j\}$ ▷ Ignore defeated arguments
      $S' \leftarrow S' \cup \{\alpha_i\}$ ▷ Construct a set of undefeated from S
    EndIf
  EndForall
  Forall $(\alpha_i, \alpha_j) \in Deft$ **Do**
    If $\alpha_i \in AR'$ *and* $\alpha_j \in S$ **Then**
      $\alpha_k = f_a(S' \cup \{\alpha_j\})$ ▷ non-deterministic selection
      $Conf' \leftarrow Conf' \cup \{conf(\alpha_k, \alpha_i)\}$
      If $\alpha_k = \alpha_j$ *and* $\neg(pref(\{\alpha_i\}, \{\alpha_j\}) \notin Pref$ *and* $pref(\{\alpha_j\}, \{\alpha_i\}) \in Pref)$ **Then**
        Select   ▷ non-deterministic selection
          $Conf' \leftarrow Conf' \setminus \{conf(\alpha_i, \alpha_j)\}$
          or
          $Pref' \leftarrow (Pref' \setminus \{pref(\{\alpha_i\}, \{\alpha_j\})\}) \cup \{pref(\{\alpha_j\}, \{\alpha_i\})\}$
        EndSelect
      EndIf
      $AR' \leftarrow AR' \setminus \{\alpha_i\}$
    EndIf
  EndForall
  Forall $\alpha_i \in AR'$ **Do**   ▷ Process non-conflicting arguments
    $\alpha_k = f_a(S)$ ▷ non-deterministic selection
    If $\{conf(\alpha_k, \alpha_i), conf(\alpha_i, \alpha_k)\} \subseteq Conf'$ **Then**
      Select   ▷ non-deterministic selection
        $Conf' \leftarrow Conf' \setminus \{conf(\alpha_i, \alpha_k)\}$
        or
        $Pref' \leftarrow Pref' \setminus \{pref(\{\alpha_i\}, \{\alpha_k\})\}$
      EndSelect
    Else
      $Conf' \leftarrow (Conf' \setminus \{conf(\alpha_i, \alpha_k)\}) \cup \{conf(\alpha_k, \alpha_i)\}$
    EndIf
  EndForall
  Return $\langle Conf', Pref' \rangle$
End

Non-deterministic selection function f_a describes a class of operator and can be replaced with a
function that select the ith enumeration of a list of arguments or choices provided from user interaction.

B

Test Cases

The following are the tables containing the test cases described in chapter 5.3.2. Each table consists of all the decision variable and values taken when determining an outcome.

Disease Site (ICD10) Code	Histology	Disease Stage			Aspect	Therapy	Evidence	Statement	Attack
		T (Tumor)	N (Node)	M (Metastasis)					
C32	M-8070/3	-	II	0	Survival	Rth	Level A	Rth (70/35/7) gives 31% 3yOS	Survival: 7
C32	M-8070/3	-	II	0	Survival	Rth + Cth	Level A	Rth (70/35/7)+Cth(Pt+5FUx3) gives 51% 3yOS	Survival: 4
C32	M-8070/3	-	II	0	Phys. Tox.	Rth ► S	Level A	Rth (70/35/7) ► S gives 70% larynx retention	PhysTox: 3
C32	M-8070/3	-	II	0	Control	Rth ► S	Level A	Rth (70/35/7) ► S gives 27% 5yDFS	Control: 3
C32	M-8070/3	-	II	0	Phys. Tox.	Rth ► S	Level A	Rth (70/35/7) ► S gives 70% toxicity	PhysTox: 1
C32	M-8070/3	-	II	0	Survival	Rth ► S	Level A	Rth (70/35/7) ► S gives 56% 5yOS	Survival: 5
C32	M-8070/3	-	II	0	Phys. Tox.	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 75% larynx retention	PhysTox: 2
C32	M-8070/3	-	II	0	Control	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 36% 5yDFS	Control: 2
C32	M-8070/3	-	II	0	Phys. Tox.	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 75% toxicity	PhysTox: 2
C32	M-8070/3	-	II	0	Survival	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 56% 5yOS	Survival: 3
C32	M-8070/3	-	II	0	Phys. Tox.	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 88% larynx retention	PhysTox: 1
C32	M-8070/3	-	II	0	Control	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 38% 5yDFS	Control: 1
C32	M-8070/3	-	II	0	Phys. Tox.	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 82% toxicity	PhysTox: 3
C32	M-8070/3	-	II	0	Survival	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 55% 5yOS	Survival: 1
C32	M-8070/3	-	II	0	Phys. Tox.	S	Level C	S (total laryngectomy) gives 0% larynx retention	PhysTox: 5
C32	M-8070/3	-	II	0	Survival	S ▷ Rth	Level A	S ▷ Rth(60/30/6) is the same as Rth(50/25/5) ▷ S	Survival: 2
C32	M-8070/3	-	II	0	Survival	S ▷ Rth	Level A	S ▷ Rth(60/30/6) is the same as Rth(70/35/7)+Cth(Pt) ► S	Survival: 1
C32	M-8070/3	-	II	0	Survival	S	Level B	S gives 40% 5yOS	Survival: 6
C32	M-8070/3	-	II	0	Psyc. Tox.	All S	Level C	S gives voice (artificial) in 80%	PsychTox: 2
C32	M-8070/3	-	II	0	Psyc. Tox.	No S	Level C	Rth gives voice (natural) in 98%	PsychTox: 1
C32	M-8070/3	-	II	0	Phys. Tox.	All Rth	Level B	Rth (IMRT) gives 39.3% xerostomia	PhysTox: 1
C32	M-8070/3	-	II	0	Phys. Tox.	All Rth	Level B	Rth (conv) gives 82.1% xerostomia	PhysTox: 4

Rth : Radiation Therapy

Cth : Chemo Therapy

S : Surgery

+ : Concurrent

► : Salvaged By

▷ : Follow by

Table B.1: Test Case A

Disease Site (ICD10) Code	Histology	Disease Stage			Aspect	Therapy	Evidence	Statement	Attack
		T (Tumor)	N (Node)	M (Metastasis)					
C32	M-8070/3	IV	-	0	Survival	Rth	Level A	Rth (70/35/7) gives 31% 3yOS	Survival: 7
C32	M-8070/3	IV	-	0	Survival	Rth + Cth	Level A	Rth (70/35/7) + Cth(Pt+5FUx3) gives 51% 3yOS	Survival: 4
C32	M-8070/3	IV	-	0	Phys. Tox.	Rth ► S	Level A	Rth (70/35/7) ► S gives 70% larynx retention	PhysTox: 3
C32	M-8070/3	IV	-	0	Control	Rth ► S	Level A	Rth (70/35/7) ► S gives 27% 5yDFS	Control: 3
C32	M-8070/3	IV	-	0	Phys. Tox.	Rth ► S	Level A	Rth (70/35/7) ► S gives 70% toxicity	PhysTox: 1
C32	M-8070/3	IV	-	0	Survival	Rth ► S	Level A	Rth (70/35/7) ► S gives 56% 5yOS	Survival: 5
C32	M-8070/3	IV	-	0	Phys. Tox.	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 75% larynx retention	PhysTox: 2
C32	M-8070/3	IV	-	0	Control	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 36% 5yDFS	Control: 2
C32	M-8070/3	IV	-	0	Phys. Tox.	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 75% toxicity	PhysTox: 2
C32	M-8070/3	IV	-	0	Survival	Cth ▷ Rth ► S	Level A	Rth (70/35/7) + neoadjCth(Pt/5FUx2) ► S gives 56% 5yOS	Survival: 3
C32	M-8070/3	IV	-	0	Phys. Tox.	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 88% larynx retention	PhysTox: 1
C32	M-8070/3	IV	-	0	Control	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 38% 5yDFS	Control: 1
C32	M-8070/3	IV	-	0	Phys. Tox.	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 82% toxicity	PhysTox: 3
C32	M-8070/3	IV	-	0	Survival	Rth + Cth ► S	Level A	Rth (70/35/7) + concCth(Pt100x3) ► S gives 55% 5yOS	Survival: 1
C32	M-8070/3	IV	-	0	Phys. Tox.	S	Level C	S (total laryngectomy) gives 0% larynx retention	PhysTox: 5
C32	M-8070/3	IV	-	0	Survival	S ▷ Rth	Level A	S ▷ Rth(60/30/6) is the same as Rth(50/25/5) ▷ S	Survival: 2
C32	M-8070/3	IV	-	0	Survival	S ▷ Rth	Level A	S ▷ Rth(60/30/6) is the same as Rth(70/35/7) + Cth(Pt) ► S	Survival: 1
C32	M-8070/3	IV	-	0	Survival	S	Level B	S gives 65% 5yOS	Survival: 2
C32	M-8070/3	IV	-	0	Psyc. Tox.	All S	Level C	S gives voice (artificial) in 80%	PsychTox: 2
C32	M-8070/3	IV	-	0	Psyc. Tox.	No S	Level C	Rth gives voice (natural) in 98%	PsychTox: 1
C32	M-8070/3	IV	-	0	Phys. Tox.	All Rth	Level B	Rth (IMRT) gives 39.3% xerostomia	PhysTox: 1
C32	M-8070/3	IV	-	0	Phys. Tox.	All Rth	Level B	Rth (conv) gives 82.1% xerostomia	PhysTox: 4

Rth : Radiation Therapy

Cth : Chemo Therapy

S : Surgery

+ : Concurrent

► : Salvaged By

▷ : Follow by

Table B.2: Test Case B

Index

Author Index